



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Signal, Image, Acoustique et Optimisation (SIAO)

Présentée et soutenue par :

Jérôme SEVERINI

Le jeudi 7 octobre 2010

Titre :

Estimation et Classification de Signaux Altimétriques

Ecole doctorale :

Mathématiques Informatique Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (IRIT)

Directeur(s) de Thèse :

Jean-Yves Tournet : Professeur E.N.S.E.E.I.H.T, Toulouse

Corinne Mailhes : Professeur E.N.S.E.E.I.H.T, Toulouse

Rapporteurs :

Nadine Martin : Directeur de Recherche CNRS GIPSA-LAB, Grenoble

Cédric Richard : Professeur Université de Nice Sophia-Antipolis

Autre(s) membre(s) du jury :

Régine André-Obrecht : Professeur Université Paul Sabatier Toulouse (Examinatrice)

Audrey Giremus : Maître de Conférences Université Bordeaux 1 (Examinatrice)

Pierre Thibaut : Ingénieur de Recherche Collecte Localisation Satellites, Toulouse (Invité)

Résumé

Les missions spatiales d’observation de la terre et des océans sont de plus en plus nombreuses (Topex/Poséidon, Envisat, Jason, Altika, ...) et de plus en plus précises dans le relevé de données. Elles permettent d’étudier la dynamique de mécanismes géophysiques à partir des mesures de plusieurs instruments embarqués. Dans cette thèse, nous nous intéressons plus particulièrement aux données altimétriques dont le principe de base consiste à mesurer la distance entre le satellite et la surface analysée en utilisant la durée de propagation d’une impulsion radar. Cependant d’autres informations pertinentes, telles que la vitesse des vents de surface et la hauteur moyenne des vagues, sont issues de l’altimétrie spatiale. Une grande majorité des mesures étant faites en haute mer, nous avons plus particulièrement étudié les relevés effectués en milieu hauturier pour lesquels une modélisation théorique des échos mesurés existe pour ce type de données.

Ce modèle a permis de développer une méthode d’estimation des paramètres d’intérêts des échos altimétriques basée sur la méthode du maximum de vraisemblance. Toutefois, l’implantation des estimateurs du maximum de vraisemblance n’est pas facile à mettre en œuvre et ce qui est aujourd’hui utilisé est une approximation de ces estimateurs. Dans ce cas précis, il ne s’agit pas d’un estimateur optimal comme nous avons pu l’illustrer dans [MTST08] ou encore comme cela est détaillé en Annexe A. C’est pourquoi nous nous sommes intéressés dans un premier temps à d’autres estimateurs afin d’en comparer les performances avec les estimateurs actuellement utilisés. En particulier, nous proposons

une méthode d'estimation bayésienne nous permettant de modéliser un certain nombre d'informations, non exploitables avec le maximum de vraisemblance, à travers des lois *a priori*. Différents types de lois *a priori* sont utilisés, et nous présentons les différences de performances de chaque estimateur [SMTT08].

Outre l'estimation des paramètres des échos altimétriques, la diversité des formes d'onde étudiées peut amener à l'utilisation d'un algorithme adapté en fonction du type de surface visualisé, ou simplement en fonction de la forme des échos eux-mêmes. Ce traitement adapté permet par la suite d'extraire des informations différentes qu'il s'agisse de signaux provenant d'océans, des déserts, des forêts, ... Ainsi, dans un deuxième temps, nous avons étudié les performances de différentes méthodes de classification permettant de différencier les signaux altimétriques suivant le type de surface observée ou suivant leur forme [TMST10b].

Enfin, nous avons considéré le cas particulier de mesures effectuées près des côtes. Ces signaux spécifiques ressemblent fortement aux formes d'onde en milieu hauturier à ceci près qu'ils sont perturbés par une composante fortement piquée. Nous proposons une nouvelle modélisation de ces échos dont nous avons validé les performances sur des traces réelles d'échos [TMST10a]. Cette dernière partie ouvre la voie à de futures études. La méthode de classification suivant la forme des échos que nous avons proposée permet dans un premier temps d'isoler les échos « classiques » pour lesquels une méthode d'estimation bayésienne permet d'obtenir des performances intéressantes. Pour les autres échos de formes différentes, il est nécessaire de proposer de nouveaux modèles. Le dernier chapitre de cette thèse s'est intéressé à un modèle pour certains échos « piqués » mais il reste d'autres formes d'échos pour lesquels le problème n'est pas résolu.

Abréviations et notations

Abréviations

ACP	Analyse en Composantes Principales
AFD, ALD, LDA	Analyse Factorielle Discriminante
BRC	Borne de Cramer-Rao
EAP	Espérance <i>A Posteriori</i>
FSR	<i>Flat Surface Response</i>
MAP	<i>Maximum A Posteriori</i>
MCMC	<i>Monte Carlo Markov Chain</i>
MLE	<i>Maximum Likelihood Estimator</i>
MPSRF	<i>Multivariate Potential Scale Reduction Factor</i>
MQE	<i>Mean Quadratic Error</i>
MSE	<i>Mean Square Error</i>
OC-SVM	<i>One Class Support Vector Machine</i>
pdf	<i>Probability Density Function</i>
PSRF	<i>Potential Scale Reduction Factor</i>
SSH	<i>Sea Surface Height</i>
SVDD	<i>Support Vector Data Description</i>
SVM	<i>Support Vector Machine</i>

Notations standard

\sim	distribué suivant
\in	appartient à
\propto	proportionnel à
\mathcal{I}	fonction indicatrice
\mathbb{R}	l'ensemble des réels
\mathbb{R}^+	l'ensemble des réels positifs
\mathbb{E}	espérance mathématiques

Notations relative à la modélisation

P_u	<i>Power Unit</i>
σ_0	P_u exprimé en décibel (dB)
τ	milieu du front de montée
sw_h	<i>Significative Wave Height</i> ou hauteur moyenne des vagues
ξ	dépointage de l'antenne
λ_s	courbure du front de montée
N_t	bruit thermique
k	numéro de porte
K	nombre de porte
L	nombre de <i>looks</i> , ou de vues
R_k	bruit multiplicatif du <i>speckle</i>
s	modèle de Hayne
y	signal observé
p	modèle du pic

Table des matières

Résumé	iii
Abréviations et notations	v
1 Altimétrie	1
1.1 L’altimétrie spatiale	1
1.2 Les principes de l’altimétrie	4
1.3 Traitement des formes d’onde en milieu hauturier	10
1.4 Perturbations spécifiques des formes d’onde	17
2 Estimation des paramètres des formes d’onde	23
2.1 Le modèle bayésien	23
2.2 Méthode de simulation de Monte Carlo	26
2.3 Contrôle de la convergence	29
2.4 Applications sur des signaux synthétiques	32
2.5 Applications sur des signaux réels	45
2.6 Cas particulier du dépointage de l’antenne	50
3 Classification des formes d’onde altimétriques	59
3.1 Introduction	59
3.2 Présentation théorique de la classification	61
3.3 Classification de différents types de surface	72

3.4	Problème de classification PISTACH	86
4	Formes d'onde altimétriques non-océaniques	109
4.1	Introduction	109
4.2	Modélisation proposée	111
4.3	Méthode du maximum de vraisemblance	113
4.4	Les performances	115
4.5	Extension à la présence de plusieurs pics	126
4.6	Conclusion	128
5	Conclusions et perspectives	131
A	Les bornes de Cramer-Rao	137
B	Méthode de Quasi-Newton	147
	Liste des publications	151
	Bibliographie	153

Table des figures

1.1	Augmentation du niveau de la mer d'environ 3mm par an depuis le début des mesures effectuées par le satellite Topex/Poseidon. © <i>University of Colorado</i>	2
1.2	Les différentes missions altimétriques. © <i>CNES/CLS</i>	3
1.3	Évolution de la précision des mesures depuis les premières missions altimétriques. © <i>CNES</i>	5
1.4	Schéma du principe de l'altimétrie. © <i>CNES</i>	6
1.5	Formation d'une forme d'onde sur un océan de surface plane. © <i>CNES</i> . . .	9
1.6	Formation d'une forme d'onde sur un océan de surface agité. © <i>CNES</i> . . .	9
1.7	Exemples de formes d'onde réelles de l'altimètre Topex.	10
1.8	Principe d'acquisition du signal altimétrique.	11
1.9	Paramètres géophysiques d'une forme d'onde théorique.	13
1.10	Pourcentage des échos types rencontrés en milieu hauturier. © [Thi08] . .	18
1.11	Formation d'une forme d'onde proche des côtes.	19
1.12	Ile de Pianosa, la flèche rouge représente la trajectoire du satellite. . . .	20
1.13	Exemple de contamination de forme d'onde lors de l'approche du satellite vers l'île de Pianosa. © [GEVQ ⁺ ar].	21
2.1	Comparaison des MSEs en fonction de σ_0 : ✱ MCMC et △ MLE. . . .	34
2.2	Comparaison des MSEs en fonction de swh : ✱ MCMC et △ MLE. . . .	35
2.3	Densités de Probabilité (i.e pdf) <i>a posteriori</i> marginales des trois paramètres. .	36

2.4	Histogrammes et interpolations des <i>priors</i> des paramètres : —* histogramme, — interpolation.	37
2.5	Comparaison des MSEs en fonction de σ_0 : —* MCMC et —△ <i>MLE</i>	38
2.6	Exemples de loi <i>a posteriori</i> . Légende : —* vraisemblance, —△ loi <i>a priori</i> , et —○ loi <i>a posteriori</i> conditionnelle.	39
2.7	Trace des paramètres « au fil de l'eau ».	40
2.8	Comparaison des MSEs en fonction de t : —* MCMC et —△ <i>MLE</i>	42
2.9	Exemple de convergence de 5 chaînes de Markov pour swh.	43
2.10	Erreur quadratique pour une réalisation —, MSE —.	44
2.11	Estimation de P_u , τ et de swh sur les échos de la trace réelle.	46
2.12	Estimation de signaux synthétiques modélisés à partir de la double convolution de Brown (en bleu) et modélisés à partir du modèle de Hayne (en vert).	47
2.13	Comparaison des échos moyennés avec l'écho théorique du modèle de Hayne.	48
2.14	Estimation de P_u , τ et de swh sur des échos de la trace réelle avec la nouvelle log-vraisemblance.	50
2.15	Influence de ξ sur une forme d'onde.	51
2.16	Comparaison des MSEs en fonction de ξ : —* MCMC et —△ <i>MLE</i>	54
2.17	Critère MPSRF de convergence en fonction du nombre d'échantillons d'intérêt.	56
2.18	Comparaison des MSEs en fonction de ξ : —* MCMC et —△ <i>MLE</i>	57
3.1	Schéma de principe pour la classification.	61
3.2	Schéma de principe pour la détection d'anomalies.	68
3.3	Représentation des signaux en fonction de X_1 et X_2	70
3.4	Illustration de l'évolution du critère d'alignement en fonction de la variance du noyau gaussien du SVM.	71
3.5	Résultats de l'utilisation des SVMs pour le critère d'alignement le plus faible.	71

3.6	Résultats de l'utilisation des SVMs pour le critère d'alignement le plus fort.	72
3.7	Exemples de formes d'onde altimétriques selon leurs types de surface. . . .	73
3.8	ACP des formes d'onde altimétriques pour des paramètres déterminés à une fréquence de 1Hz.	75
3.9	Somme cumulée des valeurs propres de l'ACP pour des paramètres obtenus à une fréquence de 1Hz.	76
3.10	Information de Fisher et somme cumulée des valeurs propres issues de l'AFD pour des paramètres obtenus à une fréquence de 1Hz.	78
3.11	AFD des formes d'onde altimétriques pour des paramètres déterminés à une fréquence de 1Hz.	79
3.12	AFD des formes d'onde altimétriques pour des paramètres déterminés à une fréquence de 1Hz dans le cas de 2 classes de Glaces.	80
3.13	ACP des formes d'onde altimétriques pour des paramètres déterminés en 18Hz.	82
3.14	Somme cumulée des valeurs propres de l'ACP pour des paramètres obtenus à une fréquence de 18Hz.	82
3.15	Information de Fisher et somme cumulée des valeurs propres issues de l'AFD.	83
3.16	AFD des formes d'onde altimétriques.	84
3.17	Schéma de principe pour la classification des formes des échos du projet PISTACH.	87
3.18	Histogramme et fonction de répartition estimée des MQE3s.	88
3.19	Histogramme et fonction de répartition estimée des MQE4s.	89
3.20	Représentations des signaux altimétriques en fonction de leurs MQEs. . . .	90
3.21	Histogrammes des paramètres de la classe normale.	91
3.22	Évolution du critère d'alignement en fonction de la variance du noyau gaussien testé à partir de 100 signaux d'apprentissage par classe.	94
3.23	Résultat de la zone de séparation obtenue avec le SVM de critère d'alignement maximum (à partir de 100 signaux d'apprentissage).	95

3.24	Évolution du critère d'alignement en fonction de la variance du noyau gaussien testé à partir de 500 signaux d'apprentissage par classe.	97
3.25	Résultat de la zone de séparation obtenue avec le SVM de critère d'alignement maximum (à partir de 500 signaux d'apprentissage).	97
3.26	Évolution du critère d'alignement en fonction de la variance du noyau gaussien testé à partir de 1000 signaux d'apprentissage par classe.	99
3.27	Résultat de la zone de séparation obtenue avec le SVM de critère d'alignement maximum (à partir de 1000 signaux d'apprentissage).	99
3.28	Critère de Fisher des variables.	103
3.29	Nouveau schéma de principe pour la classification des formes des échos. .	106
4.1	Les trois types de manifestation d'un pic dans une forme d'onde pour des signaux extraits du projet PISTACH [Thi08].	111
4.2	Échos estimés par les trois méthodes d'estimation (<i>MLE3</i> , <i>MLE3 Peak</i> et <i>Nelder-Mead</i>) d'une forme d'onde avec un pic avant le front de montée, $\theta_b = (P_u, \tau, \text{swh}) = (130, 31, 2)$ et $\theta_p = (A_1, k_1, v_1) = (400, 26, 1)$	117
4.3	Comparaison des MSEs en fonction de swh dans le cas d'un pic avant le front de montée.	118
4.4	Échos estimés par les trois méthodes d'estimation (<i>MLE3</i> , <i>MLE3 Peak</i> et <i>Nelder-Mead</i>) d'une forme d'onde avec un pic sur le front de montée, $\theta_b = (P_u, \tau, \text{swh}) = (130, 31, 2)$ et $\theta_p = (A_1, k_1, v_1) = (400, 31, 1)$	119
4.5	Comparaison des MSEs en fonction de swh dans le cas d'un pic sur le front de montée.	120
4.6	Forme d'onde avec swh = 10m et un pic sur le front de montée.	121
4.7	Échos estimés par les trois méthodes d'estimation (<i>MLE3</i> , <i>MLE3 Peak</i> et <i>Nelder-Mead</i>) d'une forme d'onde avec un pic sur le plateau final, $\theta_b = (P_u, \tau, \text{swh}) = (130, 31, 2)$ et $\theta_p = (A_1, k_1, v_1) = (400, 75, 1)$	122
4.8	Comparaison des MSEs en fonction de swh dans le cas d'un pic sur le plateau final.	123

4.9	Comparaison des estimations de formes d'onde réelles de type « Brownien ».	124
4.10	Comparaison des estimations de formes d'onde réelles avec perturbation d'un pic sur le plateau final.	125
4.11	Deux exemples d'échos perturbés par 4 pics et des δ associés.	127
4.12	Comparaison des estimations entre la méthode mono-pic et l'approche multi-pics sur deux signaux réels.	129
A.1	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les trois paramètres en fonction de L ($P_u = 160$, $\tau = 32$, $\text{swh} = 6$).	140
A.2	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les trois paramètres en fonction de P_u ($L = 90$, $\tau = 32$, $\text{swh} = 6$).	141
A.3	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les trois paramètres en fonction de τ ($L = 90$, $\text{swh} = 6$, $P_u = 160$).	141
A.4	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les trois paramètres en fonction de swh ($L = 90$, $\tau = 32$, $P_u = 160$).	142
A.5	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les quatre para- mètres en fonction de L ($P_u = 160$, $\tau = 32$, $\text{swh} = 6$, $\xi = 0.1$).	144
A.6	Comparaison entre les estimations de P_u avec un modèle de Brown à 3 et 4 paramètres ($L = 90$, $\tau = 32$, $\text{swh} = 6$, $\xi = 0.1$).	144
A.7	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les quatre para- mètres en fonction de swh ($L = 90$, $\tau = 32$, $P_u = 160$, $\xi = 0.1$).	145
A.8	Bornes de Cramer-Rao et <i>MSEs</i> du Pseudo-MLE pour les quatre para- mètres en fonction de ξ ($L = 90$, $\tau = 32$, $P_u = 160$, $\text{swh} = 6$).	145

Liste des tableaux

2.1	Potential Scale Reduction Factors.	44
2.2	Potential Scale Reduction Factors.	53
3.1	Table de confusion du classifieur bayésien en % après l'utilisation de l'ACP sur les paquets de 20 échos (calculés à la fréquence de 1Hz).	77
3.2	Table de confusion du classifieur en % après l'utilisation d'une AFD sur les paramètres des échos calculés en 1Hz.	77
3.3	Table de confusion du classifieur en % après l'utilisation d'une AFD sur les paramètres des échos calculés en 1Hz suite à la composition de 2 classes de Glaces.	79
3.4	Table de confusion du classifieur en % après l'utilisation de l'ACP écho par écho.	83
3.5	Table de confusion du classifieur en % après l'utilisation d'une AFD sur les paramètres des échos.	85
3.6	Table de confusion pour le seuillage des MQE3s.	89
3.7	Table de confusion pour le seuillage des MQE4s.	89
3.8	Table de confusion pour le seuillage des deux MQEs.	90
3.9	Table de confusion pour le seuillage des paramètres issus du <i>MLE3</i>	92
3.10	Table de confusion pour le seuillage des paramètres issus du <i>MLE4</i>	92
3.11	Table de confusion obtenue à partir du SVM avec 100 signaux d'appren- tissage par classe.	94

3.12	Table de confusion obtenue à partir du SVM avec 500 signaux d'appren-	
	tissage par classe.	96
3.13	Table de confusion obtenue à partir du SVM avec 1000 signaux d'appren-	
	tissage par classe.	98
3.14	Table de confusion moyenne obtenue avec les paramètres du <i>MLE3</i>	100
3.15	Table de confusion moyenne obtenue avec les paramètres du <i>MLE4</i>	101
3.16	Table de confusion en utilisant le classifieur bayésien avec les distances de	
	Mahalanobis sur 13 classes.	104
3.17	Table de confusion en utilisant le classifieur bayésien avec les distances de	
	Mahalanobis sur 6 regroupements.	105

Liste des algorithmes

1	Algorithme de l'échantillonneur de Gibbs.	27
2	Algorithme de Metropolis-Hastings.	28
3	Algorithme de « Metropolis within Gibbs à 3 pas de Metropolis-Hastings ».	30
4	Algorithme de « Metropolis within Gibbs à 4 pas de Metropolis-Hastings ».	52

CHAPITRE 1

Altimétrie

Sommaire

1.1	L'altimétrie spatiale	1
1.2	Les principes de l'altimétrie	4
1.3	Traitement des formes d'onde en milieu hauturier	10
1.4	Perturbations spécifiques des formes d'onde	17

1.1 L'altimétrie spatiale

1.1.1 Les objectifs de l'altimétrie

Bien qu'il soit communément acquis que la forêt Amazonienne est le « poumon » de notre planète, dans la réalité les océans s'avèrent le véritable « organe » de la Terre. Représentant près de 71% de la surface planétaire, les océans sont des mécanismes complexes régulateurs d'oxygène, de gaz carbonique mais aussi de puissants régulateurs des températures atmosphériques et donc climatiques. La compréhension des mécanismes complexes des océans est ainsi importante pour mieux préserver ces vastes espaces.

Depuis longtemps nous nous interrogeons sur le fonctionnement des océans, mais ce n'est que depuis l'apparition des techniques spatiales d'observation de la Terre qu'il est possible d'étudier ces différents mécanismes à l'échelle planétaire. L'étude océanographique couvre plusieurs disciplines telle que l'étude de la tectonique des plaques, des cycles biogéochimiques, des écosystèmes marins ou encore l'analyse des liens possibles entre l'océanographie physique (les courants marins, les vagues, les marées, ...) et les

variations du climat. La mesure de la hauteur des océans, des vents de surface (fortement liés aux températures des océans), ou encore de la hauteur des vagues sont un ensemble de paramètres nécessaires à l'étude des océans mais aussi au suivi de leurs évolutions : l'altimétrie spatiale est l'une des disciplines le permettant. À titre d'exemple, la figure 1.1 illustre l'évolution du niveau global des océans de 1993 à 2008, étude rendue possible grâce à l'altimétrie spatiale. L'ensemble des missions altimétriques et les

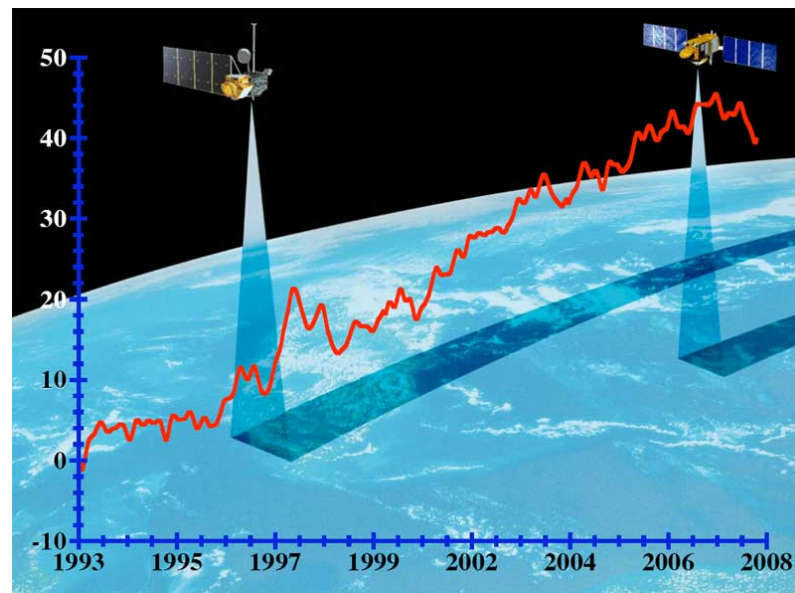


FIG. 1.1 – Augmentation du niveau de la mer d'environ 3mm par an depuis le début des mesures effectuées par le satellite Topex/Poseidon. © *University of Colorado*.

différentes avancées scientifiques en termes de modélisation et de prédiction dans l'océanographie opérationnelle permettent de prévoir l'état des océans. À l'instar des bulletins climatiques, le groupe MERCATOR proposa le premier bulletin à haute résolution des courants marins en 2005. De plus, plusieurs projets comme GMES (Global Monitoring for Environment and Security) ou GEOSS (Global Earth Observation System of Systems) intègrent l'altimétrie spatiale pour la protection et la surveillance de l'environnement en étudiant l'atmosphère, les océans ainsi que le couvert végétal.

1.1.2 Les missions altimétriques

L'altimétrie spatiale apparaît dans les années 1970 et s'impose comme un outil d'observation de la Terre. Les premières missions embarquant un altimètre sont avant tout pluridisciplinaires : GEOS-3 (1975), SEASAT (1978), GEOSAT (1985) ou ERS 1 (1991). Dans les années 1990 l'altimétrie spatiale fait l'objet d'un intérêt grandissant avec le satellite franco-américain Topex/Poséidon lancé en 1992. Il s'agit de la première mission à long terme de mesure topographique des océans qui s'achève en 2006. Ce satellite embarquait alors un altimètre américain (TOPographic EXperiment) et un altimètre européen (POSEIDON), dont les principales missions consistaient à étudier la circulation océanique, le transport de chaleur dans les océans, à modéliser les marées et surveiller les variations du niveau moyen des océans à l'échelle mondiale. L'Europe poursuit ses efforts avec les missions ERS-1 (1991) et ERS-2 (1995). De leur côté, les Etats-Unis lancèrent la mission militaire Global Follow On (GFO, en 1998) dont les données ont depuis été rendues accessibles aux civils. Dans les années 2000, une nouvelle génération d'alti-



FIG. 1.2 – Les différentes missions altimétriques. ©CNES/CLS.

mètres apparaît suite aux efforts franco-américains. Notamment en 2001 avec la mission JASON-1, successeur de la mission Topex/Poséidon, placé sur la même trajectoire afin de pérenniser les mesures et ainsi approfondir nos connaissances des différents phénomènes océanographiques. L'Europe poursuit son effort en remplaçant les missions ERS par le satellite ENVISAT (2002), dédié à l'étude de l'environnement et plus précisément au changement climatique. L'ensemble de ces satellites est présenté sur la figure 1.2.

En vue de toujours améliorer la précision et la couverture des mesures, plusieurs types d'altimètres sont à l'étude :

- l'utilisation de la bande Ka qui augmentera les résolutions verticales et spatiales car l'altimètre sera moins sensible aux perturbations ionosphériques [RLV99], [VVB⁺00],
- l'interférométrie qui à l'aide de plusieurs altimètres montés sur des « bras » permettra d'obtenir simultanément plusieurs mesures qui pourront être fusionnées [DFL⁺10],
- l'altimétrie SAR/Doppler, permettant de traiter de façon cohérente des groupes d'impulsions [Fra02], [MPRM⁺08],
- l'utilisation d'une constellation de satellites afin d'améliorer la qualité des mesures ainsi que les coûts de développement et de lancement [All06].

Toutes ces études ont une ambition commune : améliorer la précision des mesures. Cette dynamique de précision des mesures s'observe au cours des différentes missions, comme cela est illustré sur la figure 1.3. Depuis les années 1970 la marge d'erreur due à l'orbite n'a cessé de diminuer.

1.2 Les principes de l'altimétrie

1.2.1 Fonctionnement général

Bien que les océans soient référencés comme l'altitude zéro de toutes les cartes de relief, dans la réalité, ces derniers possèdent un relief en perpétuel mouvement. L'évaluation du niveau des océans, ou hauteur de la mer (Sea Surface Height, SSH), s'effectue à

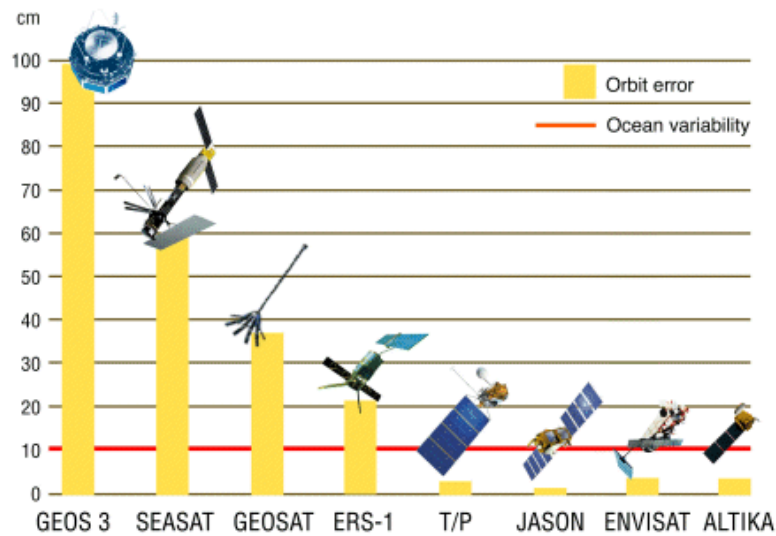


FIG. 1.3 – Évolution de la précision des mesures depuis les premières missions altimétriques. ©CNES.

partir d'une référence arbitraire que l'on nomme ellipsoïde. Cette référence correspond à la forme élémentaire de la Terre, une « sphère » aplatie aux deux pôles. La profondeur réelle des océans n'étant pas connue avec une grande précision, cette ellipsoïde de référence permet un repérage précis et homogène sur toute la planète. À partir de cette référence, le niveau de la mer SSH est considéré en théorie comme la somme d'une composante statique et d'une autre dynamique. La composante statique correspond à l'apparence de la surface des océans en absence de toutes perturbations (vents, marées, courants, etc). Il s'agit du « géoïde », qui modélise les déformations de cette surface en fonction des sous-sols marins dont les variations de masse et de densité exercent des attractions terrestres variables sur l'ensemble de la planète. La composante dynamique, quant à elle, correspond à la circulation océanique aussi appelée « topographie dynamique ». Elle comprend la circulation océanique due à la force de Coriolis (circulation liée à la rotation de la Terre), mais aussi les phénomènes variables tels que les vents ou

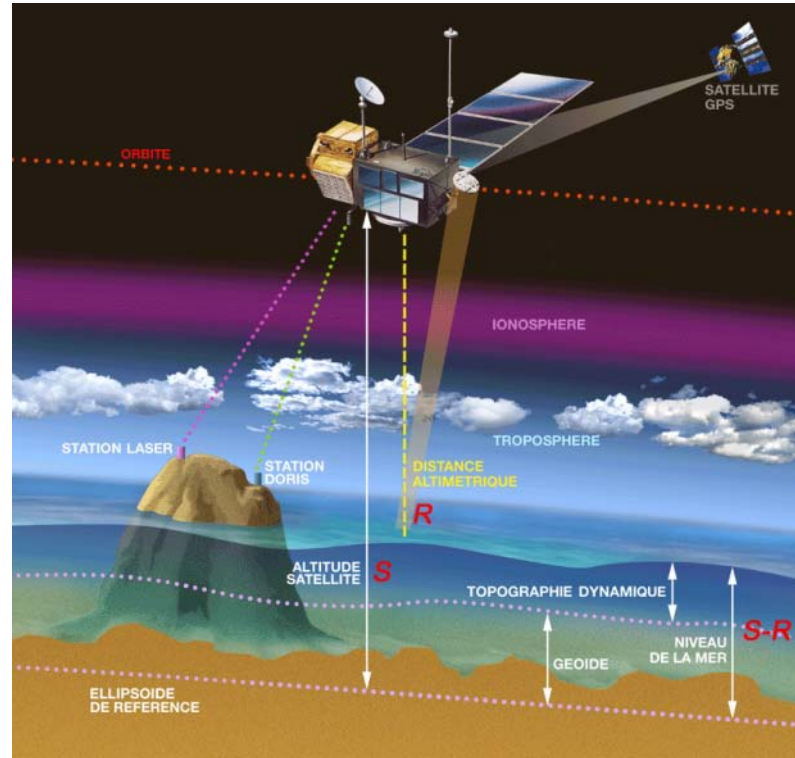


FIG. 1.4 – Schéma du principe de l'altimétrie. © CNES.

les marées fortement liés aux saisons.

$$\text{SSH} = \text{géoïde} + \text{topographie dynamique}.$$

L'accès à la topographie dynamique se fait donc en retranchant le géoïde au niveau de la mer SSH, comme illustré sur la partie du schéma 1.4, en bas, à droite. Dans la pratique, la connaissance du géoïde n'est pas parfaite et quelques modifications de l'approche sont effectuées [Oll06] pour estimer la partie dynamique. Par souci de clarté, dans les explications suivantes, ces modifications ne sont pas prises en compte.

C'est dans la mesure du niveau de la mer SSH qu'intervient l'altimétrie spatiale. Un radar altimètre est embarqué à bord du satellite pour émettre un signal à très haute fréquence en direction de la surface analysée. L'écho réfléchi par les océans permet une mesure très précise du temps de trajet t (aller et retour) entre le satellite et la surface

océanique, et permet ainsi d'en déduire la distance altimétrique R en utilisant la célérité de la lumière c :

$$R = \frac{t}{2}c.$$

Cette mesure de distance altimétrique est faite en tenant compte d'un certain nombre de corrections [CES⁺06], notamment :

- les corrections instrumentales (dues au fonctionnement de l'altimètre et au rayonnement naturel).
- les corrections atmosphériques corrigeant les allongements de trajet du signal causés par la propagation dans différentes couches de l'atmosphère. Dans la ionosphère, les électrons perturbent la vitesse de propagation, alors que dans la troposphère c'est la quantité d'eau liquide qui influe sur le signal.
- les corrections géophysiques (notamment les marées).

Le calcul du niveau de la mer SSH se fait donc en accord avec la figure 1.4 en effectuant la différence entre l'altitude du satellite notée S (calculée à partir de l'ellipsoïde) et la distance altimétrique R .

Enfin, l'évaluation précise de l'altitude du satellite se fait en employant le système *Doris*. Ce système développé par le CNES (Centre National d'Etudes Spatiales) repose sur un réseau de balises au sol émettant en direction du satellite. Une cinquantaine de balises fonctionnent actuellement à travers le monde. Un tel réseau permet alors de connaître très précisément, par effet Doppler, la vitesse du satellite sur son orbite. Il est alors possible d'en déduire la trajectoire exacte du satellite, sa position par rapport à la Terre et donc son altitude.

1.2.2 Formation d'une forme d'onde

La mesure de la distance altimétrique expliquée précédemment se fait à partir de l'écho retour réceptionné par l'altimètre. Cet écho retour est aussi appelé « forme d'onde ». La forme de l'écho altimétrique dépend directement de la surface sur laquelle l'onde radar se réfléchit. Au dessus des océans, la forme d'onde nous transmet des informations

importantes sur l'état de la mer, comme par exemple la hauteur des vagues ou la vitesse des vents de surface. Nous expliquons ici le mécanisme qui construit l'écho altimétrique.

Suite à l'émission d'une onde courte (illustrée en vert sur la figure 1.5), l'antenne radar fonctionne en mode passif en attendant la réception de l'écho réfléchi. Durant cette période, l'antenne perçoit les rayonnements naturels tels que les rayonnements telluriques, cosmiques et atmosphériques. Ces rayonnements ainsi que le fonctionnement des instruments à bord du satellite constituent le bruit thermique mesuré par l'antenne. Ainsi, avant même d'effectuer la mesure altimétrique, la puissance réceptionnée par l'antenne n'est pas nulle comme illustrée sur l'étape 1 de la figure 1.5. Lorsque l'impulsion atteint la surface observée, la surface illuminée croît alors linéairement jusqu'à obtenir une surface en forme de disque. La puissance mesurée étant directement corrélée avec la surface illuminée, croît donc linéairement, ce qui est représenté dans les étapes 2 et 3 de la figure 1.5. Dès que l'impulsion pénètre dans l'océan, la courbure de l'impulsion entraîne par la suite des surfaces illuminées en forme d'anneau de surface de plus en plus petite (étape 4 de la figure 1.5). La puissance mesurée décroît alors linéairement.

Dans le cas d'un océan agité, illustré sur la figure 1.6, le mécanisme de formation de l'écho est similaire. En effet, les premières surfaces illuminées correspondent aux crêtes des vagues (étape 2 de la figure 1.6). Au fur et à mesure que l'impulsion illumine de plus en plus les vagues, la puissance mesurée croît linéairement pour finalement atteindre un maximum lorsque les creux des vagues au nadir (plus courte distance entre le satellite et l'océan) sont illuminés. Ainsi, l'étape 1 correspond à la mesure du bruit thermique. L'étape 2 permet de mesurer la puissance réfléchie par les crêtes des vagues, tandis que l'étape 3 mesure celle des creux des vagues. Enfin, l'étape 4 représente la décroissance de la puissance réfléchie, l'impulsion pénétrant peu à peu dans l'eau. En plus de la mesure de distance entre le satellite et l'océan, les échos altimétriques permettent de déterminer la hauteur moyenne des vagues en prenant compte cette corrélation. La figure 1.7

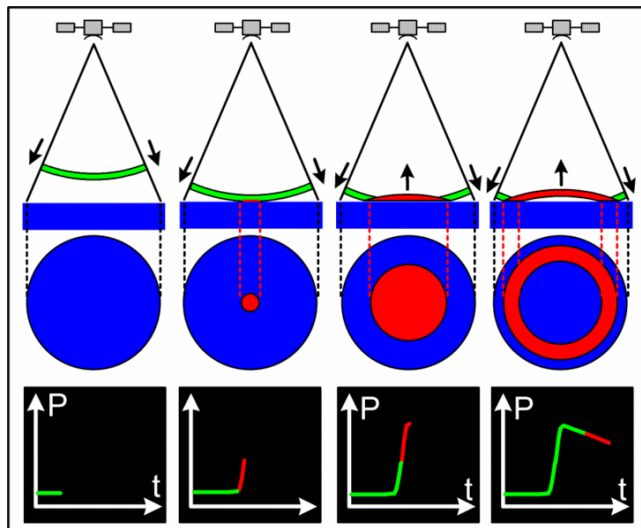


FIG. 1.5 – Formation d'une forme d'onde sur un océan de surface plane. ©CNES.

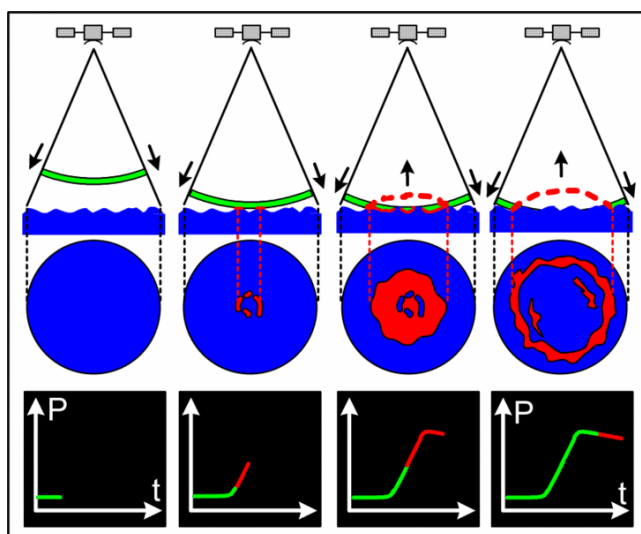
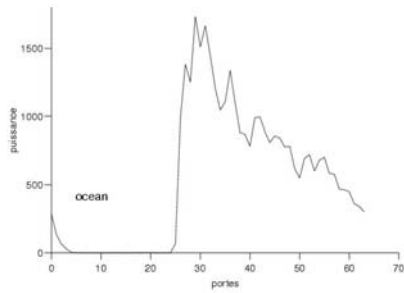
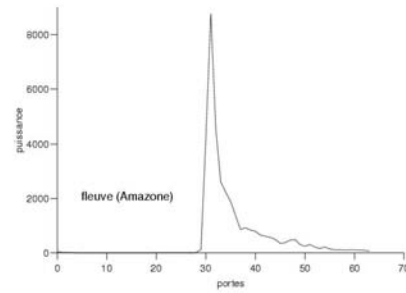


FIG. 1.6 – Formation d'une forme d'onde sur un océan de surface agité. ©CNES.

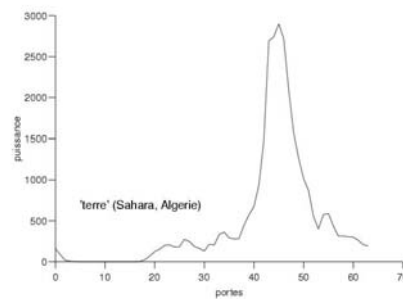
illustre trois types de formes d'onde réelles de l'altimètre Topex. La forme d'onde océanique de la figure 1.7(a) est caractéristique des échos observés de manière générale en milieu hauturier. Cette catégorie d'échos peut être décrite mathématiquement comme nous l'avons détaillé dans le paragraphe 1.3.2. Cependant, les surfaces hétérogènes telles



(a) Forme d'onde océanique



(b) Forme d'onde fluviale



(c) Forme d'onde terrestre

FIG. 1.7 – Exemples de formes d'onde réelles de l'altimètre Topex.

que les fleuves, figure 1.7(b), ou les déserts, figure 1.7(c), ne sont pas représentatives de la très grande diversité des configurations topographiques qu'il est possible de rencontrer en milieu continental, et sont donc difficilement modélisables.

1.3 Traitement des formes d'onde en milieu hauturier

1.3.1 « Tracker » et « retracking »

Dans le but de tenir compte de la topographie locale, la forme d'onde est enregistrée à l'aide d'une fenêtre d'analyse. En effet, les variations de distance altimétrique prévisibles sur une orbite donnée sont de l'ordre de 15km. Localement, cette distance peut

varier très rapidement (jusqu'à 25m/s en milieu hauturier). Afin de gérer ces variations, la fenêtre d'analyse permet d'enregistrer l'écho avec 64 ou 128 échantillons avec un pas d'environ 45cm comme cela est illustré sur la figure 1.8. Pour garantir la qualité d'enre-

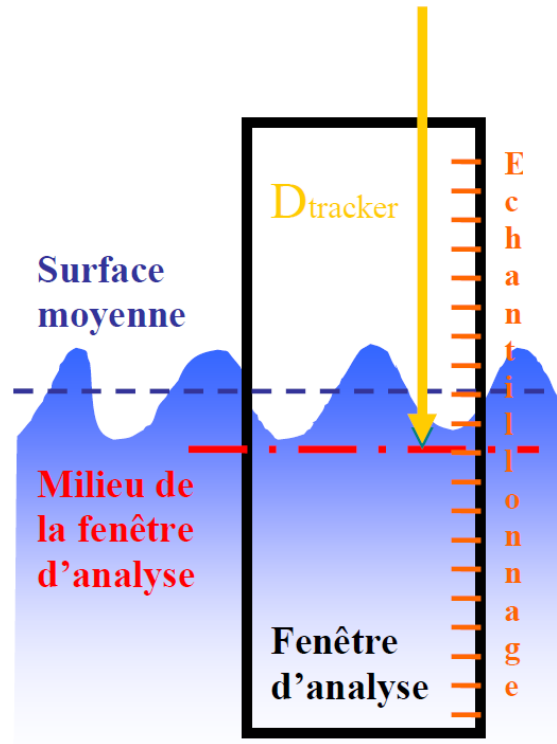


FIG. 1.8 – Principe d'acquisition du signal altimétrique.

gistement d'une mesure à l'autre, cette fenêtre est asservie en distance grâce notamment à des abbaques (méthode employée par les Américains), ou en effectuant une estimation grossière de la hauteur de la surface de la mer (méthode utilisée par les Européens). De plus, la puissance réfléchie par les surfaces analysées connaît de grandes variations, comme par exemple lors d'une transition de surface de glace vers l'eau. Il existe ici aussi un asservissement en puissance de la fenêtre afin d'éviter de saturer les composants de l'altimètre. Ces phases constituent ce qui est appelé le « *tracker* » :

- poursuite en distance ou contrôle d'altitude,
- poursuite en puissance ou contrôle automatique de gain.

La phase dite de « *retracking* » ne concerne que les missions européennes. En effet, la première estimation grossière du SSH se fait lors de la poursuite en distance du « *tracker* ». En utilisant cette estimation et la fenêtre d'analyse, la phase de « *Retracking* » consiste à effectuer une estimation de l'erreur commise et à obtenir ainsi une estimation du SSH plus fine.

1.3.2 Modélisation des formes d'onde

La forme de l'écho altimétrique observé dépend directement de la surface balayée par l'altimètre à un instant donné, comme cela est illustré sur les figures 1.5 et 1.6. Afin de modéliser le signal recueilli $s = \{s_1, \dots, s_K\}$, Brown [Bro77] propose un modèle basé sur la double convolution suivante :

$$s = FSR \otimes R_i \otimes p \quad (1.1)$$

où FSR (pour « Flat Surface Response ») est définie comme la réponse du radar à une impulsion infiniment courte envoyée sur une surface plane, R_i est la « Réponse Impulsionnelle du radar » et p est la densité de probabilité de hauteur des points spéculaires (points de la surface balayée correspondant à la plus courte distance radar-surface). C'est à partir de (1.1) que Hayne [Hay80] définit un modèle paramétrique à 5 paramètres, notés P_u , τ , swh , λ_s et ξ . La figure 1.9 permet de mieux en comprendre la signification.

Ainsi, nous distinguons trois paramètres principaux : P_u représente la puissance de l'écho, τ correspond à l'abscisse du point de puissance moitié (appelé aussi « époque du point à mi-hauteur »), tandis que swh (Sea Wave Height) est lié à la hauteur des vagues. À ces trois paramètres, s'en rajoutent deux autres : λ_s décrit la courbure du front de montée (caractéristique du skewness) et ξ correspond au dépointage de l'antenne qui influe sur la pente du plateau final. Le modèle analytique complet fixé par Hayne prenant en compte les 5 paramètres, s'écrit :

$$s_k = \frac{a_\xi P_u}{2} \exp(-v_k) \left\{ [1 + \operatorname{erf}(u_k)] + \frac{\lambda_s}{6} a_k \right\} + N_t, \quad (1.2)$$

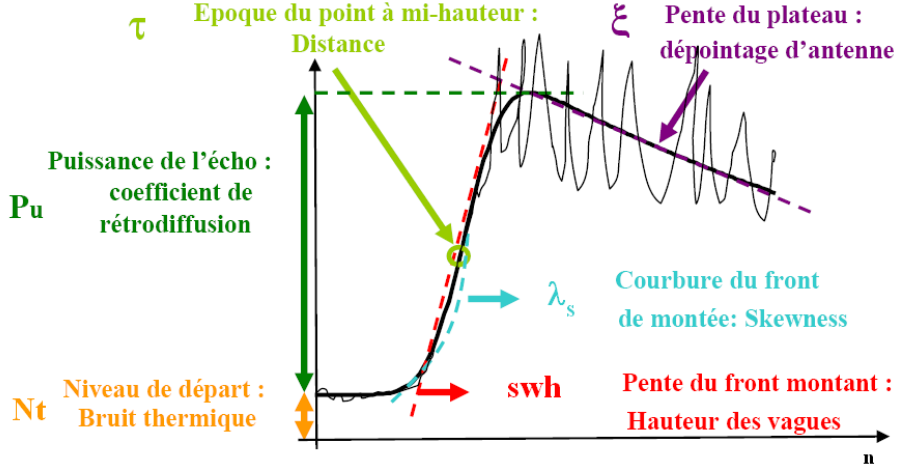


FIG. 1.9 – Paramètres géophysiques d'une forme d'onde théorique.

avec :

$$\begin{cases} a_\xi = \exp\left(-\frac{4}{\gamma} \sin^2 \xi\right), \\ a_k = \left(\frac{\sigma_s}{\sigma_c}\right)^3 \left\{ [1 + \operatorname{erf}(u_k)] \alpha^3 b_\xi^3 - \frac{\sqrt{2}}{\sqrt{\pi}} c_k \exp(-u_k^2) \right\}, \\ u_k = \frac{k - \tau - \alpha b_\xi \sigma_c^2}{\sqrt{2} \sigma_c}, \\ v_k = \alpha b_\xi \left(k - \tau - \frac{\alpha b_\xi \sigma_c^2}{2} \right). \end{cases}$$

Dans cette expression k est appelé numéro de porte ($k \in \{1, \dots, K\}$), α et γ sont des paramètres intrinsèques du radar, et N_t représente le bruit thermique. La variance σ_c^2 est reliée à la hauteur des vagues swh comme suit :

$$\sigma_c^2 = \frac{\text{swh}^2}{4c^2} + \sigma_p^2, \text{ avec } \sigma_s = \frac{\text{swh}}{2c},$$

où c est la vitesse de la lumière et σ_p un paramètre propre au radar. Enfin b_ξ et c_k sont des variables intermédiaires définies par :

$$\begin{cases} b_\xi = \cos 2\xi - \frac{\sin^2 2\xi}{\gamma}, \\ c_k = 2u_k^2 + 3\sqrt{2}\alpha b_\xi \sigma_c u_k + 3\alpha^2 b_\xi^2 \sigma_c^2 - 1. \end{cases}$$

Ce modèle peut être simplifié pour se ramener à un modèle à 4 paramètres (P_u , τ , swh , ξ) :

$$s_k = \frac{a_\xi P_u}{2} \exp(-v_k) [1 + \operatorname{erf}(u_k)] + N_t. \quad (1.3)$$

Notons que le passage du modèle simplifié à 4 paramètres (1.3) à partir du modèle (1.2) se fait en annulant le skewness ($\lambda_s = 0$) dans (1.2). De même, un modèle faisant apparaître seulement les 3 paramètres principaux (P_u , τ , swh) est aussi envisageable en supposant $\xi = 0$:

$$s_k = \frac{P_u}{2} \exp(-v_k) [1 + \operatorname{erf}(u_k)] + N_t, \quad (1.4)$$

avec :

$$\begin{cases} u_k = \frac{k - \tau - \alpha \sigma_c^2}{\sqrt{2} \sigma_c}, \\ v_k = \alpha \left(k - \tau - \frac{\alpha \sigma_c^2}{2} \right). \end{cases}$$

Notons le lien entre ce modèle simplifié (1.4) et le modèle à 4 paramètres (1.3) : en remplaçant α par αb_ξ , et P_u par $P_u a_\xi$, on retrouve le modèle (1.3). Pour commencer, nous nous concentrerons sur ce modèle simplifié afin de tester et valider la méthode d'estimation proposée.

1.3.3 Le bruitage des mesures

De par son processus de formation, une forme d'onde altimétrique est soumise à plusieurs sources de bruit. Tout d'abord le bruit thermique, dont l'origine est le fonctionnement d'appareils électriques, se modélise comme un bruit blanc gaussien spécifique au satellite. Or, une forme d'onde altimétrique s'obtient en moyennant L échos radar. Ce nombre L est aussi appelé « nombre de looks ». La moyenne des bruits thermiques associés aux L échos radar peut être considérée comme une constante propre au satellite, appelée N_t . Pour exemple, ENVISAT, un satellite basse altitude, reçoit des échos de forte puissance donnant un bruit thermique très petit face à la puissance P_u de l'écho. Ce n'est

qu'après la mise à l'échelle des données (appelée « compression ») que le bruit thermique devient réellement négligeable, voire inexistant. *A contrario*, JASON, un satellite de plus haute altitude, dispose d'échos de faible intensité pour un bruit thermique du même ordre de grandeur. N_t n'est donc plus négligeable dans la forme d'onde. La compression permet de l'atténuer mais un résidu subsiste.

L'autre source de bruit parasitant les formes d'onde est directement liée aux surfaces observées. En effet, l'observation d'un océan exposé aux rayons du soleil s'accompagne de scintillements aléatoires provoqués par la réflexion des ondes solaires. Ce phénomène se produit aussi dans le cas de mesures altimétriques. Il est à l'origine du bruit multiplicatif observé sur les mesures, appelé bruit de scintillement (ou « speckle »). Afin de réduire ce bruit, L signaux d'une même observation sont sommés de manière incohérente. En supposant la composante s_k constante sur L signaux, la résultante s'écrit :

$$y_k = \frac{s_k}{L} B_k, \quad (1.5)$$

avec

$$B_k = \sum_{l=1}^L b_{kl},$$

b_{kl} représentant l'effet du speckle pour le $l^{\text{ème}}$ signal. La variable aléatoire b_{kl} est supposée suivre une loi exponentielle de moyenne et variance unité. Une forme d'onde altimétrique est donc le résultat de la multiplication entre le modèle de Hayne s_k et le terme $\frac{B_k}{L}$, B_k suivant une loi Gamma $\mathcal{G}(L, 1)$ de moyenne $\mathbb{E}[B_k] = L$ et de variance $\text{Var}(B_k) = L$. Pour plus de précision sur les paramètres de la loi gamma, voir [RC99]. On note alors $R_k = \frac{B_k}{L}$, le bruit multiplicatif total appliqué au modèle de Hayne des formes d'onde altimétriques, la statistique de ce bruit étant une loi gamma $\mathcal{G}(L, L)$ de moyenne unité et de variance égale à $\frac{1}{L}$. Dans le cas d'un modèle de Hayne à trois paramètres (1.4), la forme d'onde altimétrique prend donc la forme suivante :

$$y_k = \left\{ \frac{P_u}{2} \exp(-v_k) [1 + \text{erf}(u_k)] + N_t \right\} R_k. \quad (1.6)$$

1.3.4 État de l'art sur l'estimation des paramètres altimétriques

À partir des données bruitées, l'estimation du vecteur paramètre $\boldsymbol{\theta} = (P_u, \tau, \text{swh})$ peut se faire à l'aide de plusieurs méthodes d'estimation :

- le maximum de vraisemblance détermine quelle valeur du vecteur paramètre maximise la statistique des observations [PCT90],
- la déconvolution de la forme d'onde consiste à extraire la densité de probabilité des hauteurs des vagues [RC89],
- les réseaux de neurones [Boe06].

La méthode actuellement utilisée, et dont les performances sont appréciables, est basée sur le maximum de vraisemblance. L'estimation par maximum de vraisemblance est une méthode courante, utilisée pour inférer les paramètres de la densité de probabilité d'un échantillon donné. Soient $\mathbf{X} = (X_1, \dots, X_K)$ un vecteur de variables aléatoires et $\mathbf{x} = (x_1, \dots, x_K)$ un vecteur d'observation de ces variables aléatoires. La densité de probabilité associée $f(\mathbf{x}|\boldsymbol{\theta})$ est supposée connue analytiquement mais l'un ou l'ensemble des attributs du vecteur paramètre $\boldsymbol{\theta}$ reste inconnu. Si les variables X_1, \dots, X_K sont indépendantes et de même loi, on obtient alors :

$$f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{k=1}^K f(x_k|\boldsymbol{\theta}). \quad (1.7)$$

La fonction $f(\mathbf{x}|\boldsymbol{\theta})$ est appelée vraisemblance des données. L'estimateur du maximum de vraisemblance $\hat{\boldsymbol{\theta}}$ est défini par :

$$\hat{\boldsymbol{\theta}} = \text{Arg max}_{\boldsymbol{\theta}} f(\mathbf{x}|\boldsymbol{\theta}). \quad (1.8)$$

$\hat{\boldsymbol{\theta}}$ est donc la valeur de $\boldsymbol{\theta}$ qui maximise la vraisemblance des données.

Dans le cas des formes d'onde altimétriques, à partir de l'équation (1.6) dans laquelle le bruit multiplicatif R_k est supposé suivre une loi $\mathcal{G}(L, L)$, la vraisemblance des données observées $\mathbf{y} = (y_1, \dots, y_K)$ s'écrit :

$$f(\mathbf{y}|\boldsymbol{\theta}) = \left[\frac{L^L}{\Gamma(L)} \right]^K \exp \left(-L \sum_{k=1}^K \frac{y_k}{s_k} \right) \left(\prod_{k=1}^K y_k \right)^{L-1} \left(\prod_{k=1}^K s_k \right)^{-L} \prod_{k=1}^K \mathcal{I}_{\mathbb{R}^+}(y_k), \quad (1.9)$$

avec $\mathcal{I}_{\mathbb{R}^+}(y_k)$ la fonction indicatrice :

$$\mathcal{I}_{\mathbb{R}^+}(y_k) = \begin{cases} 1 & \text{si } y_k \in \mathbb{R}^+, \\ 0 & \text{sinon.} \end{cases}$$

L'expression de s_k est donnée par les équations (1.4), (1.3) ou (1.2) suivant qu'on s'intéresse à un modèle à 3, 4 ou 5 paramètres. Toutefois, même en ne considérant que le modèle le plus « simple » à 3 paramètres, construire l'estimateur du maximum de vraisemblance de $\theta = (P_u, \tau, \text{swh})$ ne conduit pas à des solutions analytiques facilement exploitables. Les solutions actuellement proposées et utilisées pour l'estimation de ces paramètres correspondent à des solutions algorithmiques permettant d'approcher l'estimateur du maximum de vraisemblance.

1.4 Perturbations spécifiques des formes d'onde

En plein océan, entre 90% et 95% des formes d'onde sont de type « Brown » comme l'illustre la figure 1.10 issue d'un projet de classification des échos : PISTACH [Thi08]. Cette même figure nous permet de constater par exemple, l'influence de la proximité des côtes lorsque le satellite effectue la mesure altimétrique, la proportion de signaux dits de « Brown » diminue rapidement en fonction de cette proximité. Or, le modèle utilisé, équation (1.2), pour l'estimation des formes d'onde s'applique seulement aux échos dits de « Brown », rendant ainsi difficile l'étude de l'état de mer à proximité des côtes. Pourtant, outre la problématique globale de l'évolution du climat, l'étude des océans côtiers se révèle elle aussi importante. En effet, l'aménagement du littoral, la prévention des risques, la gestion de l'environnement et des ressources marines sont une multitude de questions auxquelles peut répondre l'étude des côtes. Les océans côtiers se composent de phénomènes complexes plus rapides et à plus petite échelle que les phénomènes en milieu hauturier. Il est donc nécessaire de mettre en place des outils et des méthodes spécifiques au traitement de ce cas particulier. Pour cela, il est avant tout nécessaire de comprendre le mécanisme de formation de ce type de forme d'onde. Comme pour la

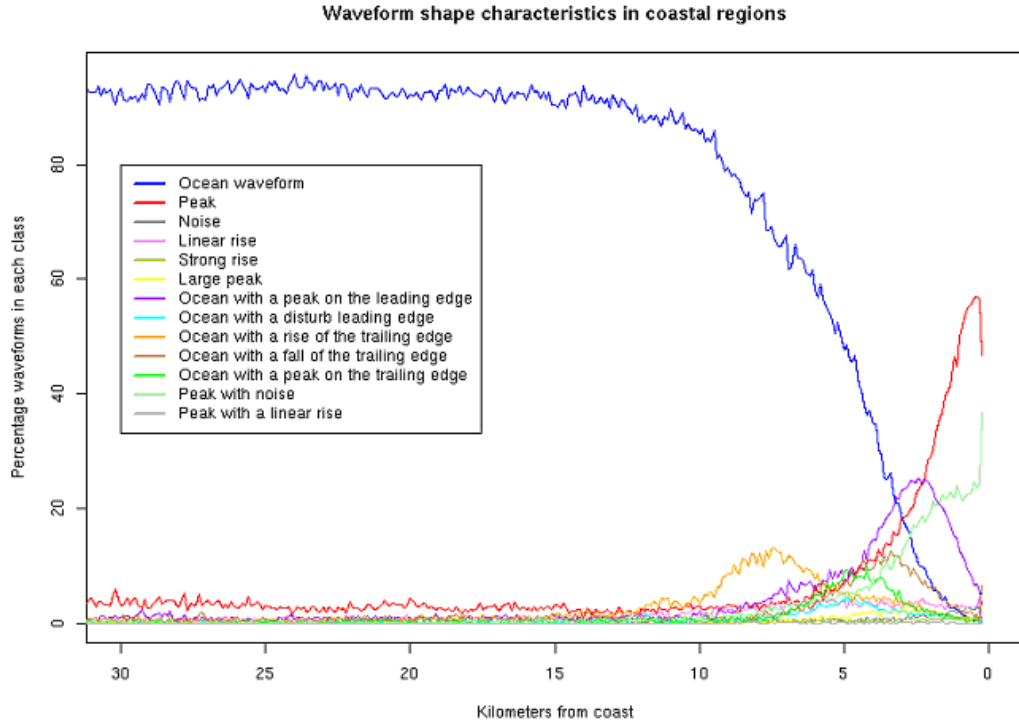


FIG. 1.10 – Pourcentage des échos types rencontrés en milieu hauturier. ©[Thi08]

formation d'une forme d'onde océanique dans la figure 1.6, les étapes de bruit thermique, de rebond de l'écho sur les crêtes des vagues et sur les creux sont identiques. Lorsque l'on considère le satellite dans la position illustrée sur la figure 1.11, vers la fin de la dernière étape de formation, les surfaces continentales reflètent à leur tour l'écho radar avec un pouvoir réfléchissant plus important. Ceci induit alors l'apparition de saut de puissance reçue par le satellite, illustré en rouge sur la figure. Plus le satellite se rapprochera de la côte, plus ce surplus de puissance se localisera en amont dans la formation de l'écho. La figure 1.13 est un exemple extrait de [GEVQ⁺ar] de mesures effectuées par le satellite aux abords de l'île de Pianosa (figure 1.12). Les abscisses représentent la latitude du satellite au moment de la mesure (et donc implicitement sa distance par rapport à l'île), les ordonnées caractérisent les portes des mesures altimétriques (i.e les échantillons) et

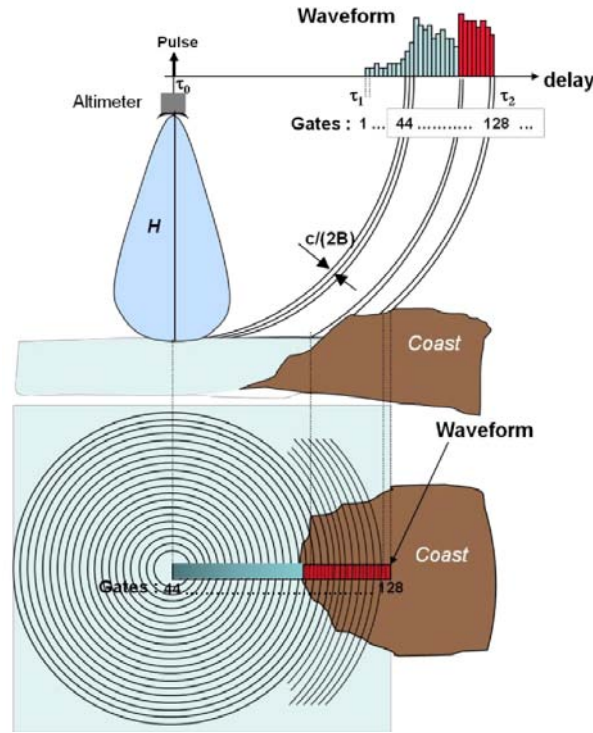


FIG. 1.11 – Formation d'une forme d'onde proche des côtes.

enfin le code de couleur correspond à des niveaux de puissance, bleu pour la plus faible et rouge pour des puissances très importantes. Sur la première partie des mesures (à gauche sur la figure 1.13), les formes d'onde sont de type « Brown » et donc peuvent être traitées par des algorithmes adaptés à ce type de forme d'onde. Cependant lorsque le satellite se rapproche de l'île, nous constatons l'apparition de sauts de puissance (en rouge) qui s'intensifient et se déplacent vers le front de montée des échos, zone délimitant la zone bleue (faible puissance) à la zone orangée, au fur et à mesure que le satellite se rapproche de l'île. Et inversement lorsque ce dernier s'en éloigne. La problématique des océans côtiers fait partie d'un ensemble de cas spécifiques de l'altimétrie océanique ([Thi08], [GEVQ⁺ar], [Bou08]). Nous pouvons aussi citer les problèmes liés aux cellules de pluie ([QGS96], [Qua97], [QSG98]), ou aux phénomènes appelés « sigmabloom » ([Gar99], [TFF07]) souvent liés à des surfaces très lisses (océan très calme avec très peu de vent)



FIG. 1.12 – Ile de Pianosa, la flèche rouge représente la trajectoire du satellite.

[TCRV06]. Ces efforts pour appliquer des traitements spécifiques aux formes d'onde en fonction des cas particuliers rencontrés, nécessitent avant tout des outils permettant automatiquement d'identifier les échos concernés pour ensuite les traiter.

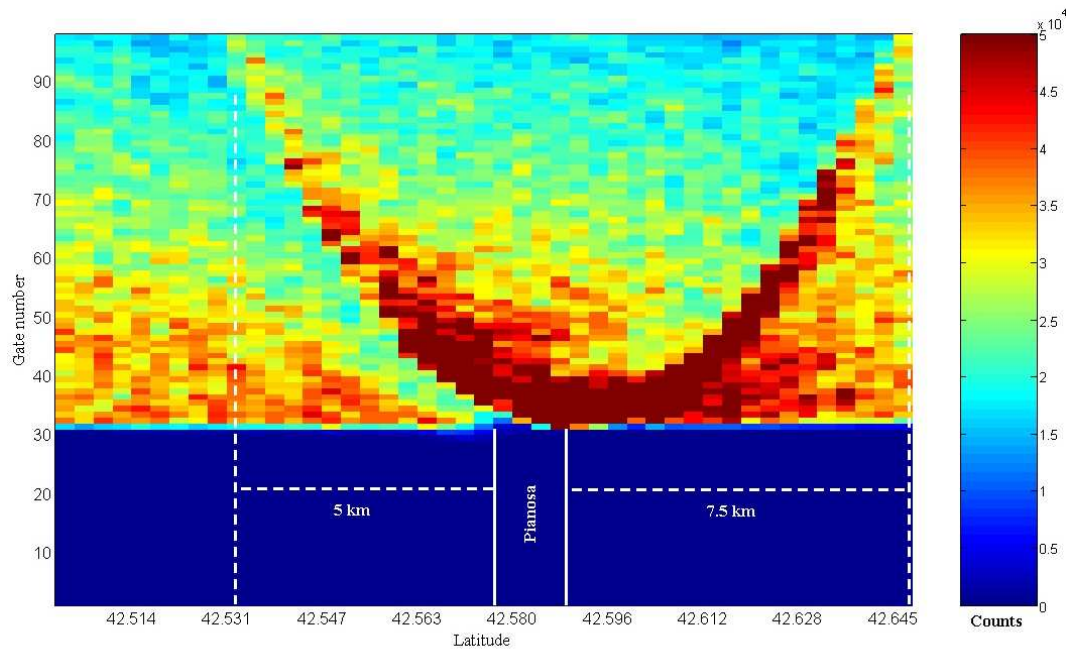


FIG. 1.13 – Exemple de contamination de forme d'onde lors de l'approche du satellite vers l'île de Pianosa. ©[GEVQ⁺ar].

CHAPITRE 2

Estimation des paramètres des formes d'onde

Sommaire

2.1	Le modèle bayésien	23
2.2	Méthode de simulation de Monte Carlo	26
2.3	Contrôle de la convergence	29
2.4	Applications sur des signaux synthétiques	32
2.5	Applications sur des signaux réels	45
2.6	Cas particulier du dépointage de l'antenne	50

2.1 Le modèle bayésien

2.1.1 Principe

On oppose couramment la conception fréquentiste et l'approche bayésienne. L'approche bayésienne permet d'inverser les probabilités en ne se limitant pas seulement à l'observation des réalisations d'un événement aléatoire. Ainsi, considérons \mathbf{X} un vecteur de variables aléatoires indépendantes d'un événement aléatoire distribué selon la loi de densité $f(\mathbf{x}|\boldsymbol{\theta})$ paramétrée par $\boldsymbol{\theta}$. Si l'on connaît les conséquences, $\mathbf{X} \sim f(\mathbf{x}|\boldsymbol{\theta})$, d'une cause (paramètre $\boldsymbol{\theta}$), l'observation des effets, réalisations de \mathbf{x} notées (x_1, \dots, x_k) , permet de remonter aux causes (paramètre $\boldsymbol{\theta}$). Ce processus se symbolise par la densité de probabilité conditionnelle $f(\boldsymbol{\theta}|\mathbf{x})$. Le théorème de Bayes [DH73] nous donne l'égalité suivante :

$$f(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{x})}. \quad (2.1)$$

Il est à noter que $f(\mathbf{x})$ est une constante de normalisation, d'où l'écriture classique :

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}), \quad (2.2)$$

où \propto signifie « proportionnelle à » et $f(\boldsymbol{\theta})$ est appelé densité de probabilité *a priori* du paramètre $\boldsymbol{\theta}$, souvent nommée *prior*. Elle résume les informations relatives à $\boldsymbol{\theta}$ indépendamment des observations. Deux cas se présentent. Dans le premier, la forme analytique de la densité $f(\boldsymbol{\theta}|\mathbf{x})$ est connue, auquel cas, l'espérance de cette loi donne une bonne estimation de $\boldsymbol{\theta}$. Dans le deuxième cas, si la forme analytique ne peut être déterminée, il est nécessaire de simuler des échantillons $\boldsymbol{\theta}^{(t)}$ distribués selon cette densité de probabilité (méthode explicitée dans la partie 2.2) afin par la suite, d'estimer $\boldsymbol{\theta}$ en utilisant soit :

- L'estimation par l'espérance *a posteriori* (EAP), simple moyennage des chaînes de Markov de densité de probabilité $f(\boldsymbol{\theta}|\mathbf{x})$, avec T le nombre d'échantillons, définie par :

$$\hat{\boldsymbol{\theta}}_{EAP} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\theta}^{(t)},$$

- L'estimation par le maximum *a posteriori* (MAP), qui optimise la probabilité du vecteur d'observation au sens du vecteur paramètre $\boldsymbol{\theta}$ prenant compte aussi du *prior* :

$$\hat{\boldsymbol{\theta}}_{MAP} = \text{Arg max}_{\boldsymbol{\theta}} f(\mathbf{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}).$$

2.1.2 Les lois *a priori*

La démarche bayésienne consiste à donner un sens à la probabilité des réalisations de $\boldsymbol{\theta}$. On parle de loi *a priori*. Cette loi, représente l'ensemble des connaissances acquises sur le vecteur paramètre $\boldsymbol{\theta}$. Dans la pratique, la loi *a priori* regroupe un grand nombre d'expertises de données. Le choix de cette loi *a priori* est bien entendu important pour la suite de l'analyse : deux lois *a priori* différentes conduisent à deux systèmes d'inférences différents, même si l'accumulation de données finit par « gommer » cette différence. Dans le cadre des signaux altimétriques, trois types de lois *a priori* sont envisagées.

Lois non informatives Tout d'abord, on peut considérer des lois *a priori* dites non informatives. Il s'agit de densités de probabilité définies avec une grande variance. Ainsi aucune partie de l'ensemble de définition des paramètres n'est statistiquement favorisée. Pour l'estimation des paramètres d'une forme d'onde, les lois *a priori* sont définies comme des lois uniformes. De ce fait, en reprenant l'équation (2.2), l'inférence bayésienne se rapproche de la méthode du maximum de vraisemblance à l'exception près que les paramètres sont définis sur un ensemble de définition plus restreint avec l'approche bayésienne.

Lois informatives Une autre approche consiste à collecter et utiliser l'historique des mesures de formes d'onde réelles. Ainsi à partir de cette collecte, il est possible d'établir un histogramme et donc d'en déduire une loi de probabilité. Afin d'en exploiter les informations apportées, nous devons déterminer la forme analytique de chacun des histogrammes. Or cela semble difficile. C'est pour cela qu'une adaptation (ou « fit ») est effectuée afin de pouvoir évaluer ces densités en tous points. L'adaptation s'est faite en utilisant une approximation de l'histogramme par une combinaison linéaire de fonctions B-spline. Il s'agit d'approcher une courbe par une concaténation de plusieurs segments polynômiaux de même ordre. Le nombre de sous-segments ainsi que l'ordre des polynômes utilisés sont à déterminer.

Lois dynamiques Enfin, il est envisageable de mettre en place des lois *a priori* qui trouvent leurs sens, non pas sur les valeurs des paramètres mais plutôt sur leurs variations au cours du temps. De manière plus précise, soit $\hat{\theta}_t$ l'estimation du paramètre θ à l'instant t . La loi *a priori* choisie à l'instant $t + 1$ est une loi normale de moyenne $\hat{\theta}_t$ et dont la variance est à définir à l'aide d'une vérité terrain.

2.1.3 La loi jointe *a posteriori*

L'estimation bayésienne se fait à partir de la loi *a posteriori*. Il s'agit d'une densité de probabilité déterminée à partir de la vraisemblance des données et des lois *a priori*

suivant l'égalité (2.2). Dans le cas de l'estimation du paramètre $\boldsymbol{\theta} = (P_u, \tau, \text{swh})$, la densité de probabilité du vecteur paramètre connaissant les données \mathbf{y} s'exprime :

$$f(P_u, \tau, \text{swh}|\mathbf{y}) \propto \exp\left(-L \sum_{k=1}^K \frac{y_k}{s_k}\right) \frac{\left(\prod_{k=1}^K y_k\right)^{L-1}}{\left(\prod_{k=1}^K s_k\right)^L} f(P_u) f(\tau) f(\text{swh}). \quad (2.3)$$

avec $f(P_u), f(\tau), f(\text{swh})$ l'expression des lois *a priori* des 3 paramètres d'une forme d'onde (dans le cas le plus simple). Dans ce contexte, quelles que soient les lois *a priori* utilisées (lois non informatives, lois informatives ou lois dynamiques), il s'avère difficile de déterminer les estimateurs bayésiens classiques (estimateur EAP et estimateur MAP) de manière analytique simple. Il est alors usuel d'utiliser des méthodes de simulation qui permettent de générer des échantillons distribués suivant la loi *a posteriori* d'intérêt. Ces échantillons sont ensuite utilisés pour estimer les paramètres inconnus du modèle, en l'occurrence les paramètres des formes d'onde altimétriques.

2.2 Méthode de simulation de Monte Carlo

Comme il a été précisé dans la partie concernant l'inférence bayésienne, l'estimation des paramètres s'effectue en utilisant la densité de probabilité *a posteriori* des paramètres inconnus. Dans ce contexte d'estimation, ces densités peuvent être analytiquement connues ou simplement simulées. Les besoins de simuler des échantillons distribués suivant une loi cible orientent notre choix vers des méthodes de Monte Carlo comme les méthodes MCMC (Markov Chain Monte Carlo).

2.2.1 Échantillonneur de Gibbs

L'échantillonneur de Gibbs est un algorithme permettant de générer des séquences d'échantillons à partir de la densité de probabilité jointe d'une ou plusieurs variable(s) aléatoire(s). C'est à partir de ces séquences qu'il est possible d'approximer les densités de probabilités conditionnelles *a posteriori*. L'échantillonneur de Gibbs est applicable lorsque

la densité de probabilité conditionnelle de chaque paramètre inconnu est connue explicitement. Les séquences ainsi générées pour chacune des variables à estimer se construisent conditionnellement aux valeurs courantes des autres variables au cours de l'inférence. L'échantillonneur de Gibbs appliqué au problème d'estimation des paramètres altimétriques est présenté dans ALGO. 1.

ALGO. 1 Algorithme de l'échantillonneur de Gibbs.

$$P_u^0 \sim f(P_u)$$

$$\tau^0 \sim f(\tau)$$

$$\text{swh}^0 \sim f(\text{swh})$$

POUR $t = 1$ à T FAIRE

$$P_u^{t+1} \sim f(P_u | \mathbf{y}, \tau^t, \text{swh}^t)$$

$$\tau^{t+1} \sim f(\tau | \mathbf{y}, P_u^{t+1}, \text{swh}^t)$$

$$\text{swh}^{t+1} \sim f(\text{swh} | \mathbf{y}, P_u^{t+1}, \tau^{t+1})$$

FIN

Pour l'ensemble des loi *a priori* dans cette étude, les densités de probabilité conditionnelles ne sont pas analytiquement connues et donc difficilement simulables. Afin de pouvoir tout de même exploiter l'algorithme de l'échantillonneur de Gibbs, nous implémentons l'algorithme de Metropolis-Hastings pour la génération d'échantillons des lois conditionnelles.

2.2.2 Metropolis-Hastings

L'algorithme de Metropolis-Hastings se fonde sur un mécanisme d'acceptation/rejet d'échantillons afin d'obtenir une chaîne de Markov ergodique de densité de probabilité correspondant à la loi cible : ici, les densités de probabilité conditionnelles *a posteriori*. Soit $f(\cdot)$ la fonction de densité de probabilité cible à simuler. On définit une loi de

proposition $q(\cdot)$, permettant de générer les différents candidats. Notons $\boldsymbol{\rho}$ la chaîne de Markov résultante de la simulation. Dans un premier temps, la séquence $\boldsymbol{\rho}$ est initialisée : $\rho^1 \sim q(\cdot)$. Ensuite un candidat ϕ est généré à partir de la loi de proposition $q(\cdot)$. On calcule alors le rapport d'acceptation/rejet afin de décider de la pertinence de l'échantillon ϕ , qui sera ou non affecté à la chaîne $\boldsymbol{\rho}$, connaissant les observations regroupées dans \mathbf{y} . Ces opérations sont alors répétées un grand nombre de fois pour finalement obtenir une chaîne de Markov qui, après convergence (ou « chauffage »), correspond à la génération d'échantillons distribués suivant la loi $f(\cdot)$. L'algorithme est détaillé dans ALGO. 2.

ALGO. 2 Algorithme de Metropolis-Hastings.

$$\rho^1 \sim q(\cdot)$$

POUR $t = 1$ à T FAIRE

$$\phi \sim q(\cdot)$$

$$\rho^{t+1} = \begin{cases} \phi \text{ avec une prob } \min \left\{ \frac{f(\phi|\mathbf{y})}{f(\rho^t|\mathbf{y})} \frac{q(\rho^t)}{q(\phi)}, 1 \right\} \\ \rho^t \text{ sinon} \end{cases}$$

FIN

Remarquons que dans le cas, où la loi *a priori* se trouve être la même que la loi de proposition, alors le rapport d'acceptation/rejet se simplifie :

$$\frac{f(\phi|\mathbf{y})}{f(\rho^t|\mathbf{y})} \frac{q(\rho^t)}{q(\phi)} = \frac{f(\mathbf{y}|\phi)f(\phi)}{f(\mathbf{y}|\rho^t)f(\rho^t)} \frac{q(\rho^t)}{q(\phi)} = \frac{f(\mathbf{y}|\phi)}{f(\mathbf{y}|\rho^t)}.$$

La probabilité que $\rho^{t+1} = \phi$ devient $\min \left\{ \frac{f(\mathbf{y}|\phi)}{f(\mathbf{y}|\rho^t)}, 1 \right\}$. On se propose donc de calculer ce rapport pour un paramètre donné, les autres étant fixés. Pour simplifier l'écriture on notera ρ la donnée ρ^t , on obtient alors à partir de l'équation (1.9), avec $\mathbf{y} = (y_1, \dots, y_K)$:

$$\frac{f(\mathbf{y}|\phi)}{f(\mathbf{y}|\rho)} = \exp \left(-L \sum_{k=1}^K \frac{y_k}{s_{\phi k}} + L \sum_{k=1}^K \frac{y_k}{s_{\rho k}} \right) \frac{\left(\prod_{k=1}^K s_{\rho k} \right)^L}{\left(\prod_{k=1}^K s_{\phi k} \right)^L}, \quad (2.4)$$

où K est le nombre d'échantillons par signal, y_k est la valeur de l'échantillon à la porte numéro k , $s_{\rho k}$ et $s_{\phi k}$ représentent la valeur de l'équation (1.4) pour les paramètres ρ et ϕ respectivement.

2.2.3 Metropolis within Gibbs

Finalement, la simulation des densités de probabilités *a posteriori* des paramètres d'une forme d'onde s'effectue en combinant les deux algorithmes présentés précédemment. ALGO. 3 décrit en détail l'algorithme de « Metropolis within Gibbs ».

2.3 Contrôle de la convergence

Un problème inhérent aux méthodes MCMC est de pouvoir déterminer si les chaînes de Markov construites ont bien convergé. Pour ce faire, construire plusieurs chaînes de Markov en parallèle, initialisées aléatoirement, permet de définir plusieurs mesures de convergence [RR98]. Nous proposons ici d'utiliser le *potential scale reduction factor* [GR07] et sa forme multivariée [BG98]. Ce critère de convergence correspond à une mesure de variance intra et extra chaîne.

2.3.1 Le Potential Scale Reduction Factor

Pour une forme d'onde altimétrique donnée \mathbf{y} , M chaînes de Markov $\boldsymbol{\rho}$ de taille N et d'initialisations différentes, sont générées pour l'estimation d'un même paramètre ($\boldsymbol{\rho}_{P_u}$, $\boldsymbol{\rho}_\tau$ ou $\boldsymbol{\rho}_{\text{swh}}$). Dans ce cas précis, le critère ne se calcule que pour un seul paramètre d'intérêt (P_u , τ ou swh). La variance extra-chaîne B et la variance intra-chaîne W se définissent par :

$$B = \frac{N}{M-1} \sum_{j=1}^M (\bar{\rho}_j - \bar{\rho})^2, \quad (2.5)$$

$$W = \frac{1}{M(N-1)} \sum_{j=1}^M \sum_{t=1}^N (\rho_j^t - \bar{\rho}_j)^2, \quad (2.6)$$

ALGO. 3 Algorithme de « Metropolis within Gibbs à 3 pas de Metropolis-Hastings ».

$$P_u^0 \sim q(P_u)$$

$$\tau^0 \sim q(\tau)$$

$$\text{swh}^0 \sim q(\text{swh})$$

POUR $t = 1$ à T FAIRE

$$\phi_{P_u} \sim q(P_u)$$

$$P_u^{t+1} = \begin{cases} \phi_{P_u} \sim \min \left\{ \frac{f(\mathbf{y}|\phi_{P_u}, \tau^t, \text{swh}^t, N_t^t) f(\phi_{P_u})}{f(\mathbf{y}|P_u^t, \tau^t, \text{swh}^t, N_t^t) f(P_u^t)} \frac{q(P_u^t)}{q(\phi_{P_u})}, 1 \right\} \\ P_u^t \text{ sinon} \end{cases}$$

$$\phi_\tau \sim q(\tau)$$

$$\tau^{t+1} = \begin{cases} \phi_\tau \sim \min \left\{ \frac{f(\mathbf{y}|P_u^{t+1}, \phi_\tau, \text{swh}^t, N_t^t) f(\phi_\tau)}{f(\mathbf{y}|P_u^{t+1}, \tau^t, \text{swh}^t, N_t^t) f(\tau^t)} \frac{q(\tau^t)}{q(\phi_\tau)}, 1 \right\} \\ \tau^t \text{ sinon} \end{cases}$$

$$\phi_{\text{swh}} \sim q(\text{swh})$$

$$\text{swh}^{t+1} = \begin{cases} \phi_{\text{swh}} \sim \min \left\{ \frac{f(\mathbf{y}|P_u^{t+1}, \tau^{t+1}, \phi_{\text{swh}}, N_t^t) f(\phi_{\text{swh}})}{f(\mathbf{y}|P_u^{t+1}, \tau^{t+1}, \text{swh}^t, N_t^t) f(\text{swh}^t)} \frac{q(\text{swh}^t)}{q(\phi_{\text{swh}})}, 1 \right\} \\ \text{swh}^t \text{ sinon} \end{cases}$$

FIN

avec $\bar{\rho}_j$ défini comme la moyenne des échantillons de la $j^{\text{ème}}$ chaîne, et $\bar{\rho}$ la moyenne des $\bar{\rho}_j$:

$$\begin{cases} \bar{\rho}_j = \frac{1}{N} \sum_{t=1}^N \rho_j^t, \\ \bar{\rho} = \frac{1}{M} \sum_{j=1}^M \bar{\rho}_j. \end{cases}$$

À partir des variances intra-chaîne W et extra-chaîne B , il est possible d'obtenir un estimateur $\widehat{\text{Var}}(\boldsymbol{\rho})$ de la variance des chaînes de Markov [GR07]. Il se définit par :

$$\widehat{\text{Var}}(\boldsymbol{\rho}) = \frac{N-1}{N} W + \frac{B}{N}. \quad (2.7)$$

Le *potential scale reduction factor* (*PSRF*) s'écrit alors :

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}(\boldsymbol{\rho})}{W}}. \quad (2.8)$$

Il s'agit donc d'un rapport entre un estimateur de la variance des chaînes de Markov et de la moyenne des variances des M chaînes étudiées. En pratique un *PSRF* inférieur à 1.2 indique une bonne convergence [GCSR95].

2.3.2 L'approche multivariée

L'extension de ce critère de convergence pour des paramètres multivariés à P composantes, se fait en utilisant les matrices de variance-covariance [BG98]. Si l'on s'intéresse à l'estimation du vecteur $\boldsymbol{\rho}$, l'observation $\boldsymbol{\rho}_j^t$ fait référence au vecteur paramètre $\boldsymbol{\rho}$ (de dimension $P \times 1$) de la chaîne j à l'instant t . De manière analogue au *PSRF*, des matrices de variance-covariance intra-chaîne \mathbf{W} et de variance-covariance extra-chaîne \mathbf{B} se définissent comme suit :

$$\mathbf{B} = \frac{N}{M-1} \sum_{j=1}^M (\bar{\boldsymbol{\rho}}_j - \bar{\boldsymbol{\rho}})(\bar{\boldsymbol{\rho}}_j - \bar{\boldsymbol{\rho}})^T, \quad (2.9)$$

$$\mathbf{W} = \frac{1}{M(N-1)} \sum_{j=1}^M \sum_{t=1}^N (\boldsymbol{\rho}_j^t - \bar{\boldsymbol{\rho}}_j)(\boldsymbol{\rho}_j^t - \bar{\boldsymbol{\rho}}_j)^T, \quad (2.10)$$

où $\boldsymbol{\rho}$ est la matrice de dimension $P \times N \times M$, avec \mathbf{W} et \mathbf{B} des matrices P dimensionnelles.

Il en découle alors l'estimation de la variance-covariance *a posteriori* $\hat{\mathbf{V}}$:

$$\hat{\mathbf{V}} = \frac{N-1}{N} \mathbf{W} + \frac{\mathbf{B}}{N}. \quad (2.11)$$

Il est alors possible d'étudier la convergence d'un échantillonneur avec :

$$\hat{\mathbf{R}} = \hat{\mathbf{V}} \mathbf{W}^{-1}, \quad (2.12)$$

et de conclure lorsque $\hat{\mathbf{R}}$ est proche de la matrice identité P dimensionnelle \mathbf{I}_P . On parle de *multivariate potential scale reduction factor (MPSRF)*. Plusieurs mesures de distances entre ces deux matrices sont possibles, cependant une méthode simple et suffisante pour obtenir le critère de convergence sous forme de scalaire est définie par [BG98] :

$$\hat{R} = \sqrt{\frac{1}{P} \sum_{i,j} \hat{R}_{ij}^2}. \quad (2.13)$$

Il s'agit bien du critère P dimensionnel analogue au *PSRF*, donnant le même résultat pour $P = 1$. Ainsi l'étude de la convergence s'effectue comme dans le cas unidimensionnel.

2.4 Applications sur des signaux synthétiques

Les différents résultats de simulation de cette partie, comparent les résultats entre la méthode de référence, le *MLE* approché, algorithme proposé par la société CLS [Dum85], et l'approche bayésienne. Dans un premier temps, le modèle simplifié (1.4), avec $\boldsymbol{\theta} = (P_u, \tau, \text{swh})$, des signaux altimétriques est utilisé. Bien que le bruit thermique soit un des paramètres du modèle de Hayne, il est généralement estimé en moyennant les 5 premières portes du signal. L'algorithme de simulation de type « Metropolis within Gibbs » nécessite un temps de « chauffe » pour garantir la convergence des chaînes de Markov. Afin de valider dans un premier temps l'approche bayésienne, un grand nombre d'itérations sont associées à cette période de chauffage (« burn-in ») et à l'estimation des paramètres du modèle. L'ALGO. 3 est caractérisé par : $N = Nbi + Nbr$ avec $Nbi = 1500$ échantillons pour le « burn-in » et $Nbr = 6500$ échantillons d'intérêt pour les estimations.

2.4.1 Sans *a priori* sur les paramètres

Lois *a priori* Tout d'abord, les simulations sont menées en utilisant des lois *a priori* uniformes pour les 3 paramètres. Ces lois *a priori* sont utilisées comme lois de proposition nécessaires à l'inférence bayésienne. Les intervalles de définition de chacune des lois sont fixés à partir d'histogrammes issus d'une vérité terrain. Plus précisément, ces lois *a priori* sont définies comme suit :

$$\begin{cases} \sigma_0 \sim \mathcal{U}[9.5, 25], \\ \tau \sim \mathcal{U}[21, 42], \\ \text{swh} \sim \mathcal{U}[0, 11] \end{cases} \quad (2.14)$$

avec $P_u = 10^{\frac{\sigma_0 - k}{10}}$ où k caractérise un facteur d'échelle et σ_0 un coefficient de rétrodiffusion.

Simulations Afin de juger des performances de l'approche bayésienne, nous la comparons à l'estimateur actuellement utilisé par CLS. L'outil statistique prenant compte du biais et de la variance des estimateurs est l'erreur quadratique moyenne (MSE) : $MSE(\hat{\theta}) = \text{biais}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$.

Cette erreur quadratique moyenne est calculée à l'aide de 100 simulations de Monte Carlo pour différentes valeurs de σ_0 (ou P_u selon les simulations) et/ou de swh . Les résultats sont présentés sur les figures 2.1 et 2.2. En observant les figures 2.1 et 2.2 il est possible de remarquer que pour les paramètres σ_0 (implicitement P_u) et τ , les deux méthodes d'estimation ont des résultats similaires. Cependant, l'approche bayésienne s'avère plus précise pour l'estimation de swh (la hauteur significative des vagues). Les différences de précisions minimale et maximale entre les deux méthodes se définissent par :

$$\begin{cases} \delta_{\text{Min}}(\hat{\theta}) = \sqrt{\text{Min} [MSE_{MLE}(\hat{\theta}) - MSE_{MCMC}(\hat{\theta})]}, \\ \delta_{\text{Max}}(\hat{\theta}) = \sqrt{\text{Max} [MSE_{MLE}(\hat{\theta}) - MSE_{MCMC}(\hat{\theta})]}. \end{cases} \quad (2.15)$$

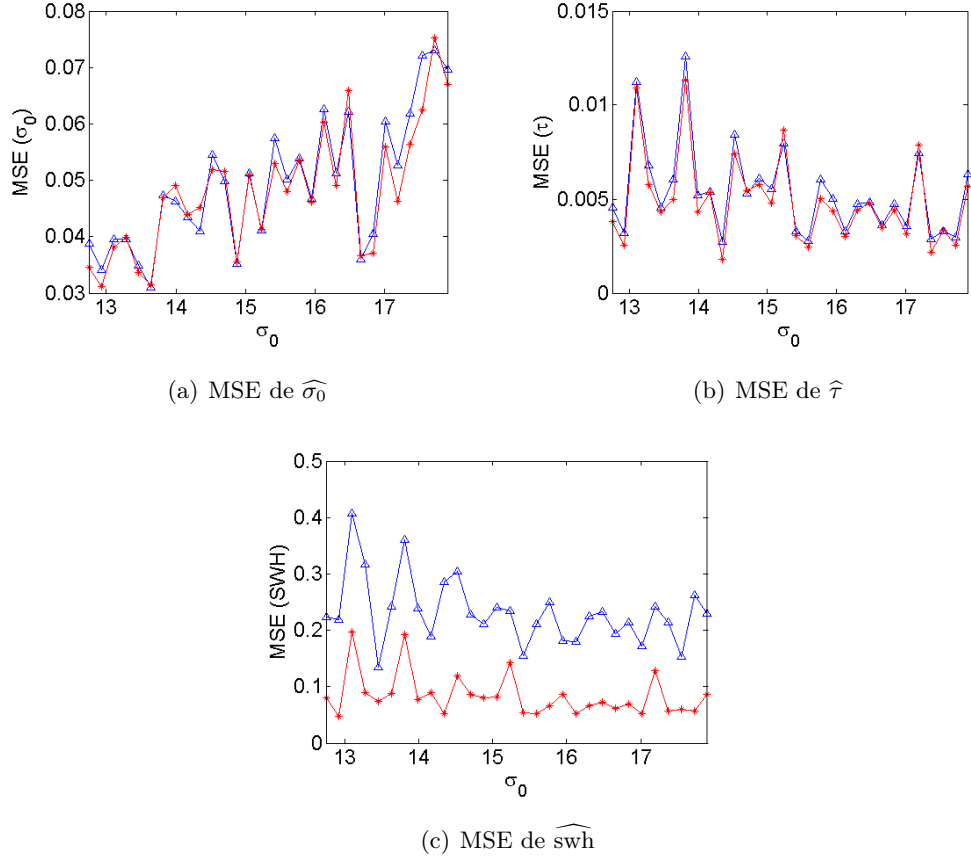
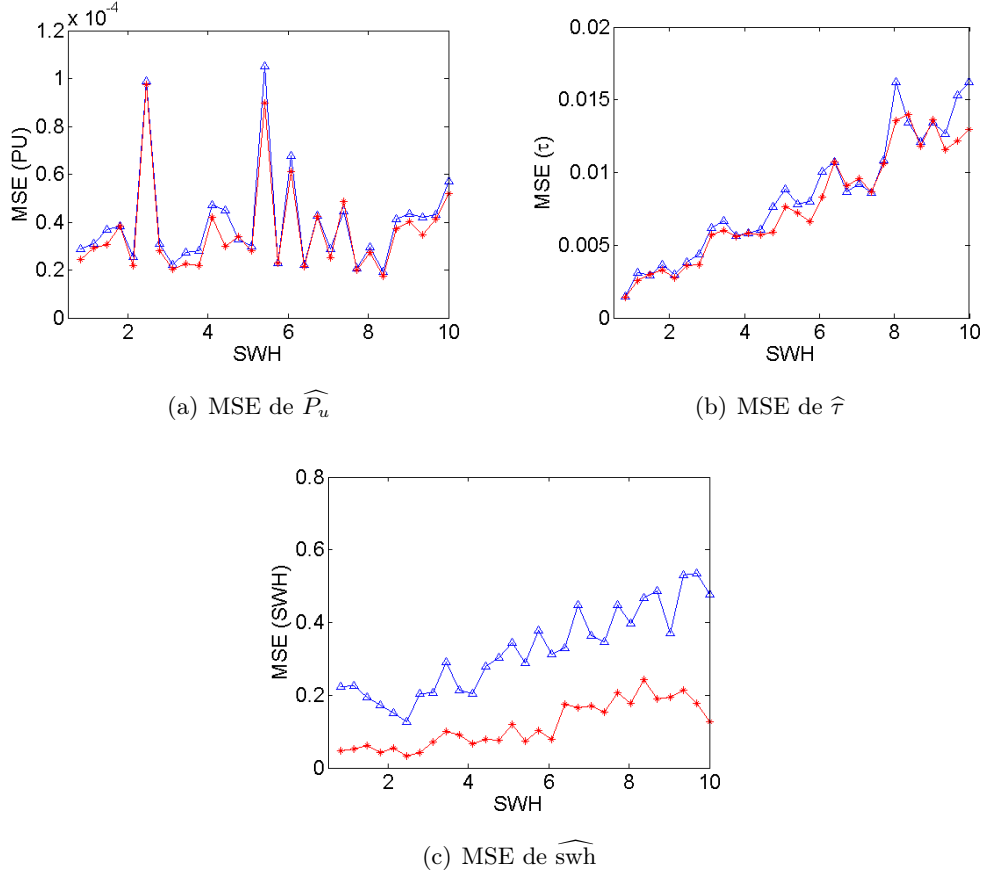


FIG. 2.1 – Comparaison des MSEs en fonction de σ_0 : $\text{---}\ast\text{---}$ MCMC et $\text{---}\triangle\text{---}$ MLE.

Dans cet exemple, $\delta_{\text{Min}}(\widehat{\text{swh}}) = 0.25\text{m}$ et $\delta_{\text{Max}}(\widehat{\text{swh}}) = 0.45\text{m}$. Cette différence s'explique par l'utilisation d'une version approchée du maximum de vraisemblance dans le MLE, due à un modèle des formes d'onde à forte non linéarité.

En plus d'une meilleure précision pour l'estimation, une des particularités de la méthode bayésienne, est qu'elle permet d'estimer la loi *a posteriori* des paramètres d'intérêts, et des intervalles de confiance associés aux estimateurs. Par exemple, la figure 2.3 montre les histogrammes des paramètres σ_0 , τ et swh générés par la méthode d'échantillonnage. Ces histogrammes sont en accord avec les vraies valeurs des paramètres pour $\theta = (\sigma_0, \tau, \text{swh}) = (13.3, 31.4, 5.25)$.

FIG. 2.2 – Comparaison des MSEs en fonction de swh : $\text{--}*$ MCMC et $\text{--}\triangle$ MLE.

2.4.2 Avec *a priori* issue de réalité terrain

Lois *a priori* Comme cela a déjà été mentionné, la méthode bayésienne utilise la notion de loi *a priori*. Or en utilisant des lois uniformes, l'information apportée au cours de l'inférence est minime. Il est intéressant d'utiliser les histogrammes issus d'un cycle, pour les paramètres P_u, τ et swh . Notons qu'un cycle représente l'ensemble des mesures satellitaires associées à un tour du globe.

Afin d'avoir des expressions analytiques des lois *a priori* concernant les divers paramètres, nous avons effectué une interpolation de ces histogrammes à l'aide de fonctions

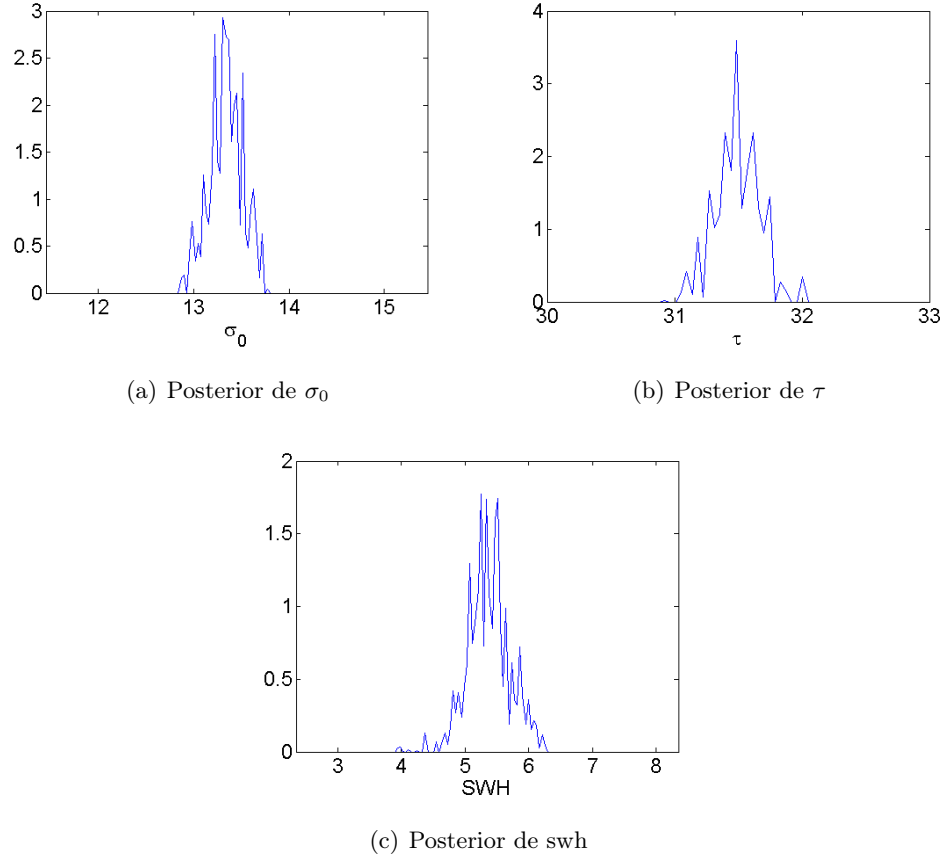


FIG. 2.3 – Densités de Probabilité (i.e pdf) *a posteriori* marginales des trois paramètres.

splines (concaténation de plusieurs segments polynômiaux de même ordre). Le nombre de sous-segments ainsi que l'ordre des polynômes utilisés sont à déterminer. Enfin, pour le milieu de front de montée τ , l'adaptation de cet histogramme s'est faite ici en utilisant une simple interpolation linéaire car l'utilisation des splines dans ce cas précis autorise des probabilités négatives. C'est-à-dire que la courbe approchante est déterminée de sorte que l'erreur quadratique entre les deux courbes (l'histogramme et l'approchante) soit minimale. La figure 2.4 illustre les histogrammes et leurs interpolations. Pour les besoins de l'inférence, les lois *a priori* de P_u (et donc σ_0) et swh sont aussi utilisées comme loi de proposition. La loi *a priori* de τ étant trop piquée, la loi de proposition définie pour

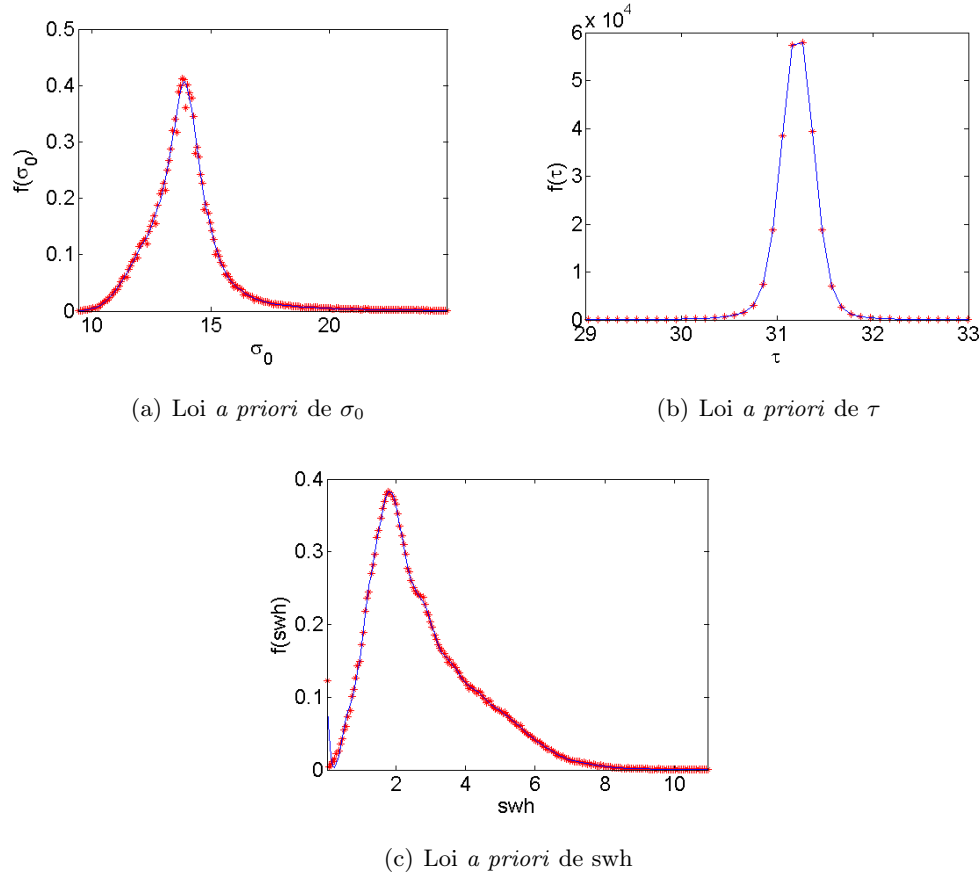


FIG. 2.4 – Histogrammes et interpolations des *priors* des paramètres : *— histogramme, — interpolation.

ce paramètre est une loi gaussienne approchant au mieux la loi *a priori*, mais dont la variance est multipliée par 5. En procédant ainsi, la loi de proposition autorise la génération de candidats considérés comme très peu probables lorsqu'on considère uniquement la loi *a priori* d'origine.

Simulations De la même manière que pour les lois *a priori* uniformes, les deux approches sont comparées à partir des MSEs des différentes estimations. Le principal problème de cette partie vient du fait que la loi *a priori* de τ est très fortement piquée. Deux types de simulation ont donc été menés.

Tout d'abord, pour des valeurs croissantes de σ_0 et générant aléatoirement swl sur son ensemble de définition (2.14), τ est généré sur un espace restreint, correspondant aux valeurs les plus probables de sa loi *a priori*. La figure 2.5(a) met en évidence que l'approche bayésienne permet d'augmenter la précision de l'estimation dans ce cas précis. Les performances pour les deux autres paramètres sont identiques à celles observées dans le cas de loi *a priori* uniformes. Cependant, pour des valeurs de τ peu probables (issue de la loi *a priori*), la figure 2.5(b) met en avant un problème dans l'inférence bayésienne. Pour mieux comprendre ces résultats, nous allons détailler un cas pour une valeur de τ

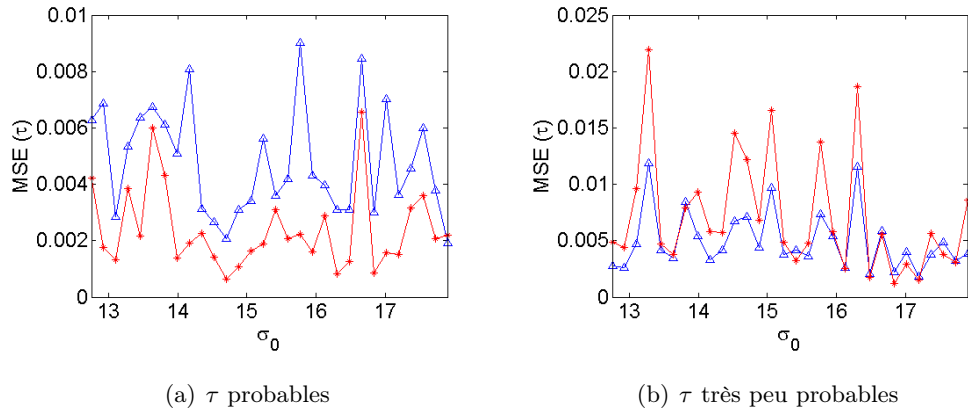


FIG. 2.5 – Comparaison des MSEs en fonction de σ_0 : * MCMC et \triangle MLE

probable, un autre pour une valeur de τ très peu probable. La figure 2.6(a) permet de visualiser, pour $\tau = 31.5$ (valeur probable), la vraisemblance des données, la loi *a priori* de ce paramètre ainsi que sa loi *a posteriori* conditionnelle estimée à partir des échantillons générés par l'algorithme de « Metropolis within Gibbs ». On remarque que la combinaison de la loi *a priori* et de la vraisemblance conduit à une loi *a posteriori* conditionnelle

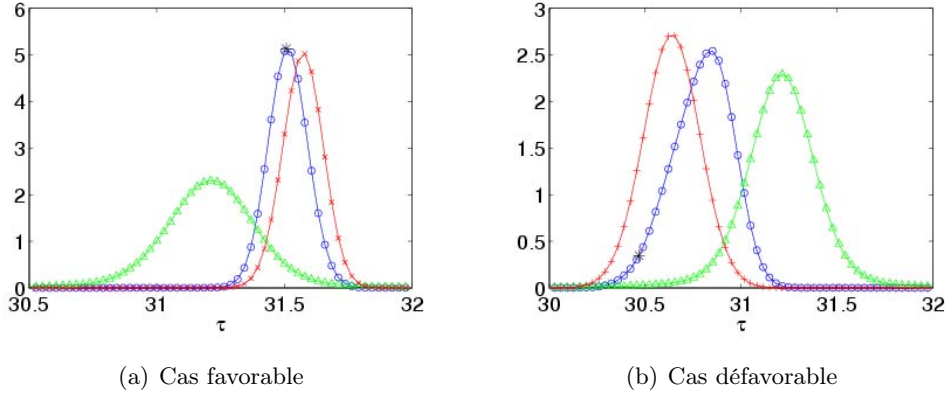


FIG. 2.6 – Exemples de loi *a posteriori*. Légende : *-* vraisemblance, \triangle loi *a priori*, et \circ loi *a posteriori* conditionnelle.

bien centrée sur la valeur de $\tau = 31.5$, valeur qu'on cherche à estimer. Dans ce cas, l'estimation bayésienne présente visiblement un avantage : l'estimateur du maximum de vraisemblance revient à chercher la valeur de τ qui maximise la vraisemblance, ce qui fournit une valeur de τ autour de 31.6, tandis qu'un estimateur bayésien (MAP par exemple), permet une estimation de ce paramètre beaucoup plus proche de la réalité. Cependant, en analysant la figure 2.6(b), on se rend compte, que dans le cas d'une valeur peu probable de τ (ici $\tau = 30.5$), l'estimateur du maximum de vraisemblance conduit à une meilleure estimation que celle obtenue par une méthode bayésienne. En effet si la vraisemblance est pratiquement centrée sur la vraie valeur $\tau = 30.5$, la loi *a posteriori* obtenue avec la loi *a priori* choisie surestime le paramètre τ .

2.4.3 Avec *a priori* sur les variations des paramètres

Lois *a priori* Afin de prendre en compte une évolution des paramètres au cours du temps, nous envisageons d'utiliser des lois *a priori* dites dynamiques. Plutôt que d'utiliser des lois *a priori* fixes sur les valeurs que peuvent prendre les différents paramètres, le but de ces lois est de mettre en évidence l'évolution des valeurs prises par les paramètres

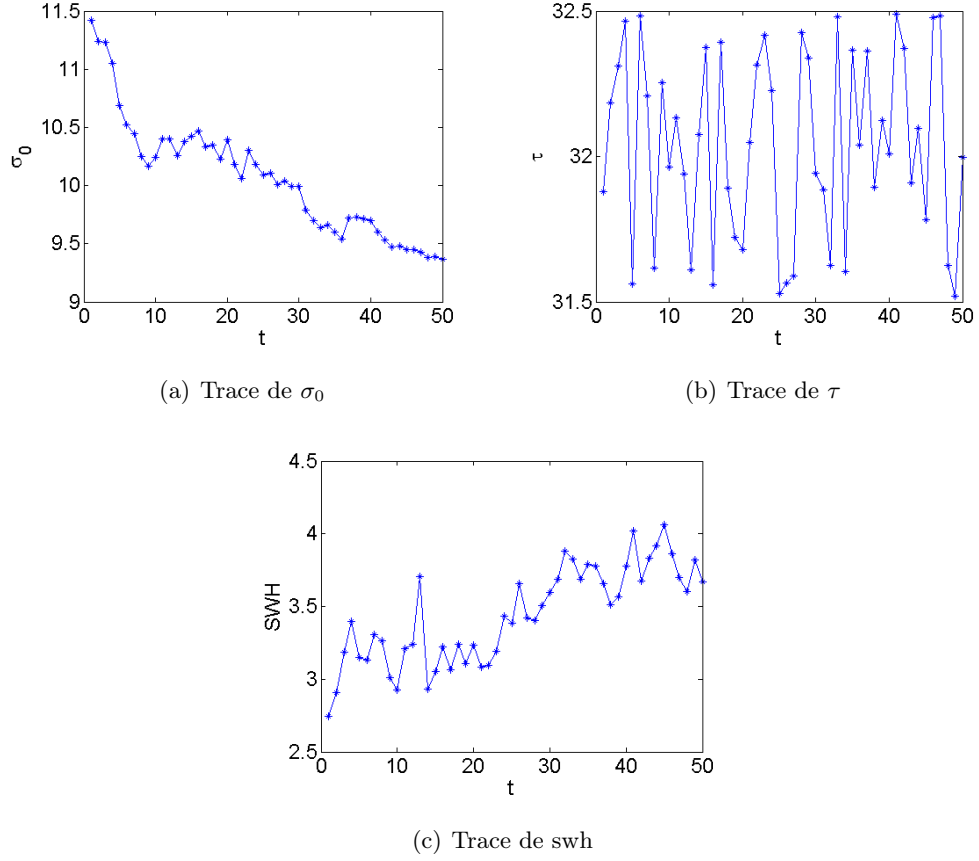


FIG. 2.7 – Trace des paramètres « au fil de l'eau ».

au cours du temps. Pour cela, les différentes simulations ne se font plus sur des valeurs quelconques des paramètres, mais « au fil de l'eau ». C'est-à-dire que les simulations reproduisent les paramètres des formes d'onde rencontrées au cours d'un cycle par un satellite. Pour cela nous disposons de traces pour σ_0 et swh. Le paramètre τ reste généré aléatoirement. Ces traces sont illustrées sur la figure 2.7. Ainsi, nous pouvons procéder à des estimations en fonction du temps t . L'intérêt de cette procédure est de pouvoir mettre en place des lois *a priori* dynamiques. Ainsi, nous proposons de définir la loi *a priori* du paramètre θ à l'instant $t + 1$ comme une loi gaussienne de moyenne θ^t et de variance fixe. Pour nos simulations, les variances choisies pour tous les paramètres sont

égales à 0.5 (cette valeur est choisie de manière heuristique mais peut être adaptée à l'aide d'un plus grand nombre d'observations des formes d'onde). L'initialisation des lois *a priori* se fait en utilisant les résultats du *MLE* approché (proposé par la société CLS) à l'instant $t = 0$. Comme cela a été fait dans les premières simulations, les lois *a priori* sont aussi utilisées comme loi de proposition pour l'inférence bayésienne.

Simulations Comme les simulations précédentes, la comparaison entre le *MLE* et l'approche bayésienne se fait en utilisant les MSEs. Pour cela, 100 estimations dynamiques, sur des formes d'onde soumises à différentes réalisations du speckle, sont effectuées afin de calculer les erreurs quadratiques moyennes (i.e les MSEs). Nous pouvons observer sur la figure 2.8 une légère amélioration des estimations par l'approche bayésienne des paramètres σ_0 et τ . En effet à partir de (2.15) il est possible de déterminer que les estimées de τ par la méthode bayésienne dynamique, sont de 1cm à 6cm plus précises qu'avec le *MLE* approché. Pour la hauteur significative des vagues, sw_h , les améliorations apportées par la nouvelle méthode permettent d'obtenir des estimées de 30cm à 60cm plus précises qu'avec le *MLE* approché.

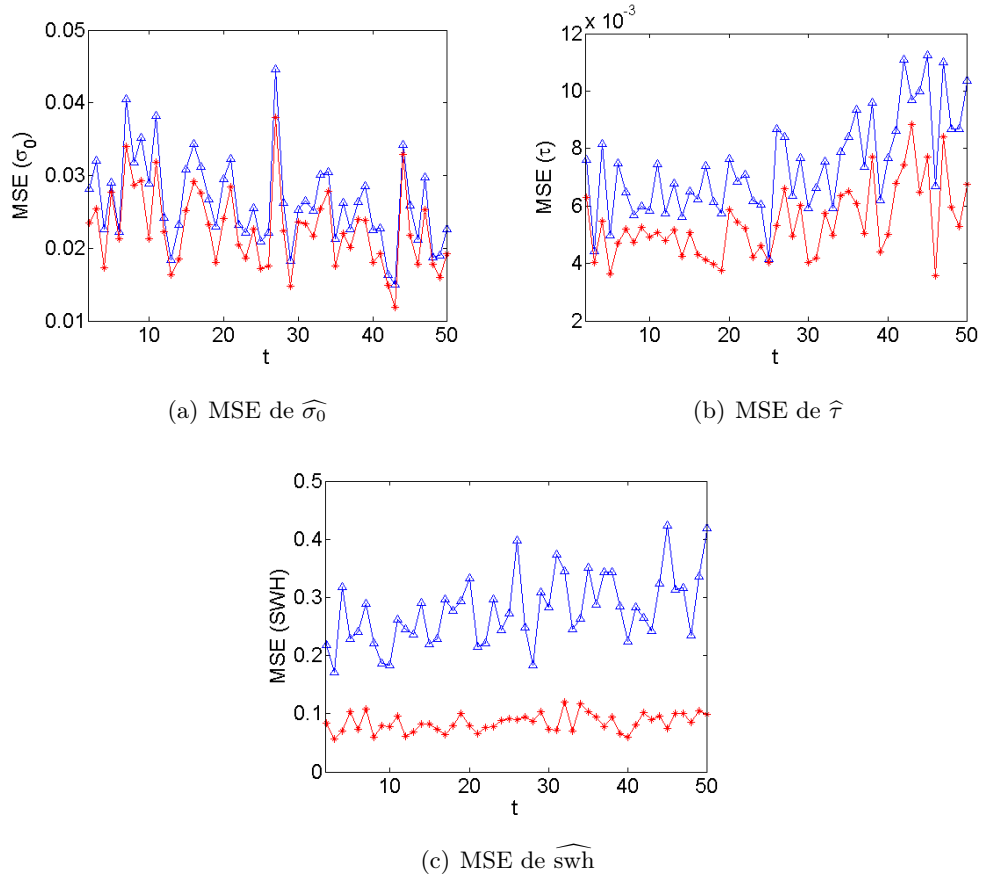


FIG. 2.8 – Comparaison des MSEs en fonction de t : $\textcolor{red}{*}$ MCMC et $\textcolor{blue}{\triangle}$ MLE.

2.4.4 Contrôle de la convergence

Comme il a été précisé dans le paragraphe 2.3, afin de garantir que l'algorithme de Metropolis within Gibbs génère bien des échantillons suivant les lois cibles, nous mesurons le *PSRF*, équation (2.8). Pour cela, $M = 100$ chaînes de Markov initialisées aléatoirement sont générées. Les conditions d'expérimentation sont identiques à celles utilisées pour les simulations précédentes : $N = Nbi + Nbr$ avec $Nbi = 1500$ échantillons pour le « burn-in » et $Nbr = 6500$ échantillons d'intérêt pour les estimations. La figure 2.9 illustre 5 chaînes de Markov pour le paramètre $swh = 6m$. Visuellement, il est possible de remarquer sur cette figure que dès les premières itérations, l'algorithme converge. Les valeurs du critère de Gelman & Rubin \hat{R} , en utilisant les Nbr dernières itérations, sont données dans le tableau 2.1. Ces valeurs confirment la convergence de l'échantillonneur puisqu'un maximum de 1.2 est conseillé pour assurer la convergence [GCSR95]. Il est donc possible d'affirmer que les chaînes de Markov ont convergé. Mais

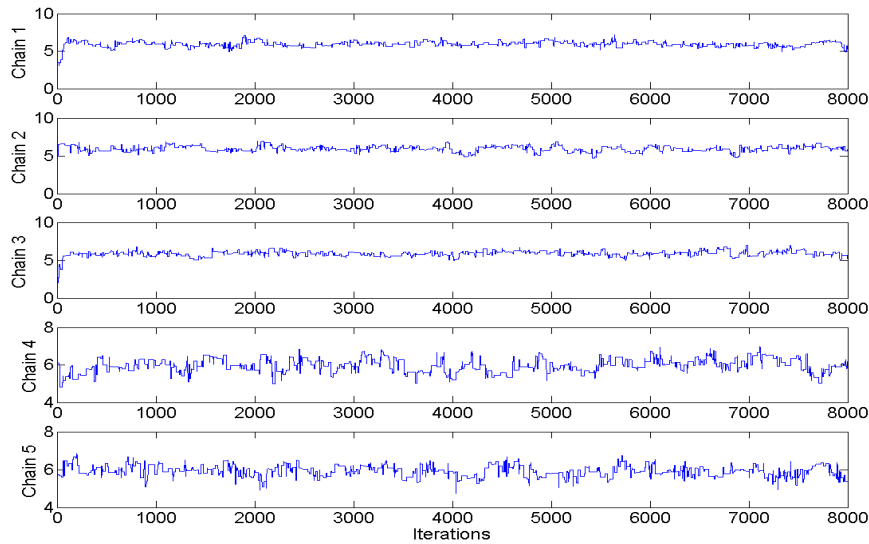


FIG. 2.9 – Exemple de convergence de 5 chaînes de Markov pour swh .

l'estimateur construit à partir des Nbr dernières itérations est-il efficace ? Pour vérifier

θ	P_u	τ	swh	$MPSRF$
\widehat{R}	1.0055	1.0187	1.0131	1.017

TAB. 2.1 – Potential Scale Reduction Factors.

que cela est bien le cas, un estimateur de référence est mis en place avec des chaînes de Markov générées avec un « burn-in » $Nbi = 1500$ et $Nbr = 100000$ échantillons. Ensuite, pour la même forme d'onde, 100 chaînes de Markov sont générées avec le même temps de chauffage et le même nombre d'échantillons d'intérêt. La figure 2.10 met en avant les erreurs quadratiques moyennes (MSEs) observées pour l'estimateur de référence et les 100 estimateurs en faisant varier Nbr pour les 100 chaînes. Il est alors possible de considérer que $Nbr = 6500$ est un bon compromis entre précision d'estimation et coût calculatoire.

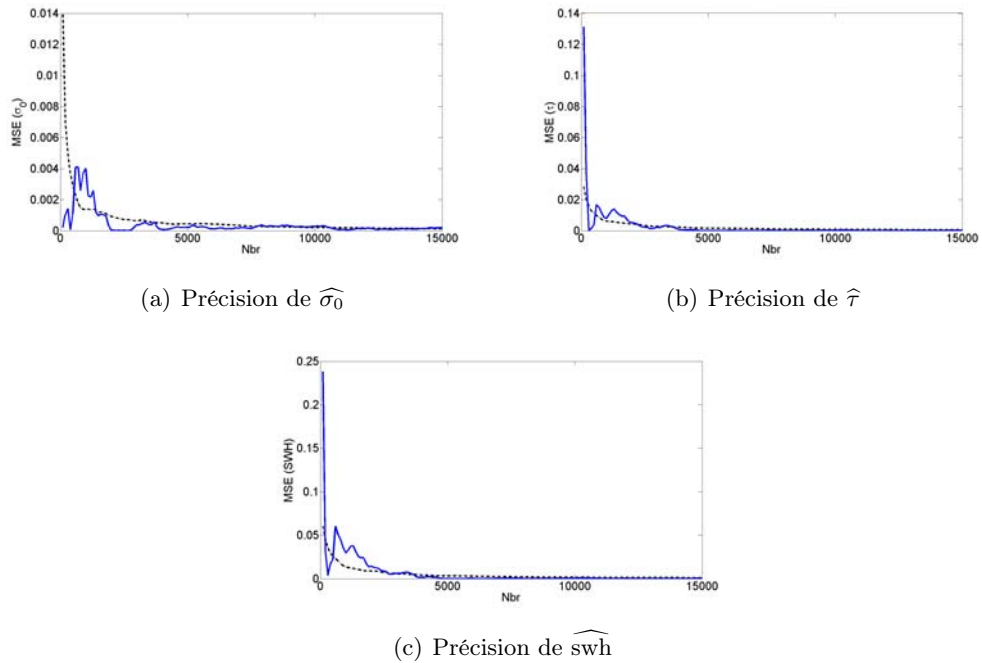


FIG. 2.10 – Erreur quadratique pour une réalisation —, MSE - - .

2.5 Applications sur des signaux réels

Dans cette partie nous nous intéressons à la comparaison des estimations obtenues par l'estimation bayésienne présentée précédemment et le *MLE* approché couramment utilisé sur des signaux réels provenant du satellite Jason-1. Pour cela nous utilisons une trace de 588 formes d'onde altimétriques pour lesquelles nous ne disposons pas de vérité « terrain », le *MLE* approché est donc utilisé comme algorithme de référence.

2.5.1 Une estimation biaisée

La comparaison des performances des deux algorithmes, sur des signaux synthétiques suivant le modèle de Hayne (1.6), nous a permis de mettre en évidence l'intérêt de l'approche bayésienne dans le cas de lois *a priori* dynamiques. En effet, en terme de variance d'estimation, l'algorithme bayésien s'avère plus efficace que le *MLE* approché. Nous cherchons donc ici à valider cette amélioration sur des signaux réels. Ne pouvant calculer des erreurs quadratiques moyennes (absence de vérité « terrain »), nous comparerons directement les estimations obtenues par les deux algorithmes. De plus, nous utiliserons un estimateur du maximum de vraisemblance « exact » au travers d'une routine d'optimisation de Nelder-Mead [NM65]. Rappelons en effet que le *MLE* approché actuellement implanté utilise des simplifications des dérivées partielles pour le calcul du gradient de la fonction cible.

La figure 2.11 permet de comparer les estimations des paramètres P_u , τ et sw_h fournies par le *MLE* approché (le plus couramment utilisé), celles obtenues avec l'approche bayésienne et l'estimateur du maximum de vraisemblance « exact » (Nelder-Mead) sur une trace de signaux réels. Pour l'estimation de P_u les trois algorithmes sont comparables. Cependant, pour les estimations de τ et sw_h , nous observons clairement un biais assez important le long de la trace. Dans cet exemple, la méthode bayésienne et Nelder-Mead estiment des hauteurs de vagues moyennes de l'ordre de 3m au lieu des 1.5m déterminés par le *MLE* approché. Ainsi, sur des signaux réels, l'estimation bayésienne et l'estimation

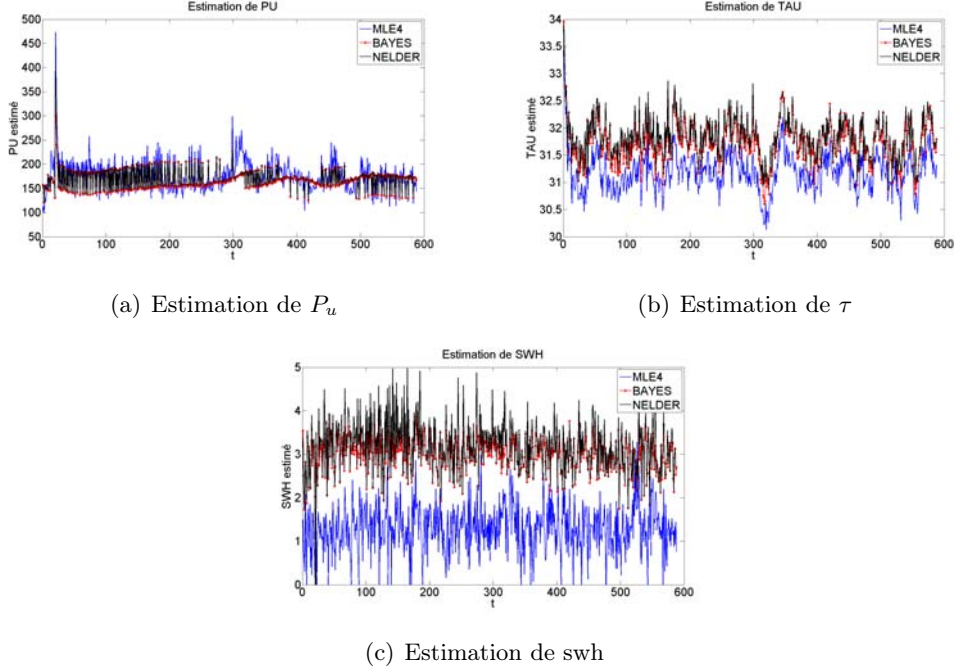


FIG. 2.11 – Estimation de P_u , τ et de swh sur les échos de la trace réelle.

du maximum de vraisemblance « exact » fournissent les mêmes résultats, pourtant ils ne sont pas en accord avec le *MLE* approché qui reste l'estimateur de référence.

2.5.2 Origine du biais

Cette différence de performance entre l'approche bayésienne (et la méthode de Nelder-Mead) et le *MLE* approché n'avait jamais été observée sur des signaux synthétiques. Afin d'identifier le problème, nous avons analysé plusieurs paramètres non pris en compte pour des signaux synthétiques détaillés dans [TMS10], tels que :

- l'estimation du nombre de look L comme par exemple en imagerie radar,
- l'influence sur les méthodes d'estimation d'un système de compression non uniforme sur un écho donné,
- l'influence sur les méthodes d'estimation d'une quantification d'un écho,
- et enfin, un problème de modélisation des données.

Nous exposons ici seulement l'analyse faite sur la modélisation. Pour cela, nous disposons de deux jeux de 100 formes d'onde fournies par la société CLS. Pour un même vecteur paramètre $\theta = (P_u, \tau, \text{swh}) = (160, 31, 2)$, le premier ensemble d'échos synthétiques est généré à partir de la double convolution de Brown [Bro77], et le deuxième à partir du modèle simplifié de Hayne [Hay80]. La figure 2.12 présente la comparaison des résultats d'estimation obtenus sur ces deux jeux de données. Par souci de clarté, nous ne présentons que les résultats obtenus avec l'algorithme du maximum de vraisemblance « exact » (i.e. approche de Nelder-Mead), sachant que l'estimateur bayésien donne les mêmes résultats et que l'estimateur du *MLE* approché conduit à des estimations non biaisées autour des valeurs attendues des paramètres. À partir de la figure 2.12, nous constatons que dans le

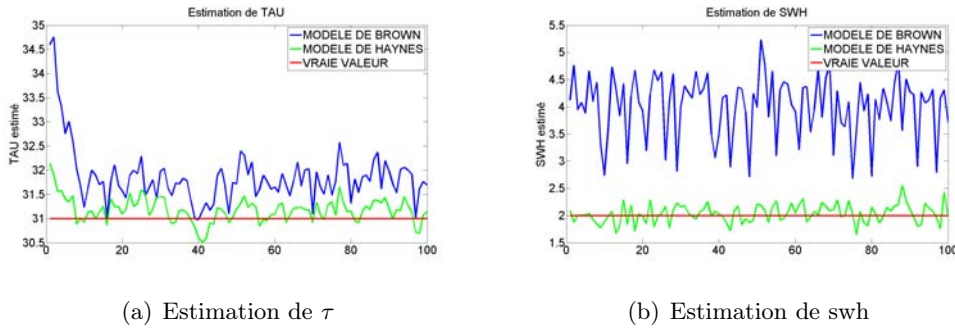


FIG. 2.12 – Estimation de signaux synthétiques modélisés à partir de la double convolution de Brown (en bleu) et modélisés à partir du modèle de Hayne (en vert).

cas de données simulées suivant le modèle de Hayne, donc sans convolution, l'estimation de τ est autour de la porte 31 et l'estimation de swh autour de 2m, confirmant le bon comportement de l'algorithme (par extension l'algorithme bayésien) sur de telles données. Cependant, dans le cas de formes d'onde simulées selon la double convolution, modèle le plus proche des échos réels, nous pouvons remarquer que les estimations de τ et swh sont biaisées. Le biais observé vient donc de la modélisation des données, plus spécifiquement des simplifications effectuées pour la mise en place du modèle de Hayne. Mais alors pourquoi le *MLE* n'est-il pas biaisé ?

Afin d'observer ces différences de modélisation, nous moyennons les échos des deux ensembles de formes d'onde altimétriques à notre disposition afin d'atténuer les effets du bruit multiplicatif. La figure 2.13 montre les deux échos moyennés obtenus ainsi que la courbe théorique du modèle de Hayne. Nous constatons que l'écho moyenné à partir des données simulées suivant le modèle de Hayne est pratiquement confondu avec l'écho théorique. En revanche, on s'aperçoit que l'écho moyenné issu d'une simulation incluant la convolution est différent, en particulier sur les portes précédant le front de montée, ce qui correspond à la courbure du front de montée. On comprend bien alors qu'un

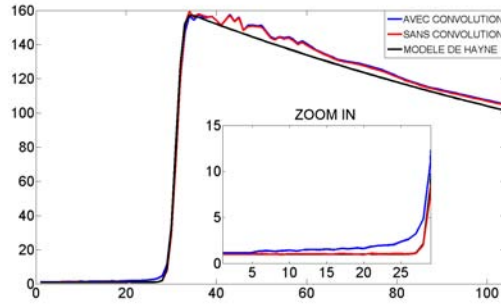


FIG. 2.13 – Comparaison des échos moyennés avec l'écho théorique du modèle de Hayne.

maximum de vraisemblance exact (et donc la méthode bayésienne) sur ce type de donnée ne soit pas performant du fait d'une vraisemblance imprécise sur une partie de l'écho. Pourtant le *MLE* approché reste insensible à cette imperfection de la vraisemblance. Cela s'explique par une simplification dans le calcul du gradient de la fonction coût qu'il cherche à minimiser comme cela est expliqué dans l'annexe B. En effet, dans la pratique les équations (B.5) et (B.6) sont remplacées par :

$$D_k = \frac{s_k - y_k}{P}, \quad (2.16)$$

$$B_{ik} = \frac{\partial s_k}{P \partial \theta_{b,i}}, \quad (2.17)$$

avec $P = 160$. Cette approximation revient à mettre un poids sur les premières portes de façon à ce qu'elles n'interviennent pratiquement pas dans le calcul du gradient et de

la hessienne de la fonction coût. Ce qui explique pourquoi le *MLE* approché n'est pas sensible aux portes précédant le front de montée (i.e correspondant à la courbure du front de montée) et permet d'obtenir de bonnes estimations sur des signaux simulés selon les deux modèles ainsi que sur les signaux réels.

2.5.3 Adaptations et résultats sur des signaux réels

Dans le cas des signaux réels, nous décidons alors d'adapter la vraisemblance des données afin de supprimer le biais observé juste avant. La log-vraisemblance actuelle d'un écho altimétrique est de la forme :

$$\ln [f(\mathbf{y}|\boldsymbol{\theta}_b)] = \sum_{k=1}^K \ln \left[\frac{L^{KL}}{(L-1)!} \right] + (L-1) \ln(y_k) - L \ln(s_k) - L \frac{y_k}{s_k}. \quad (2.18)$$

La simplification apportée par le *MLE* approché, équations (2.16) et (2.17), nous permet de définir la nouvelle vraisemblance :

$$\ln [f(\mathbf{y}|\boldsymbol{\theta}_b)] = \sum_{k=1}^K \ln \left[\frac{L^{KL}}{(L-1)!} \right] + (L-1) \ln(y_k) - L \frac{s_k^2}{2P^2} + L \frac{y_k}{P^2} s_k. \quad (2.19)$$

Nous avons donc modifié l'algorithme bayésien en introduisant cette nouvelle log-vraisemblance.

La figure 2.14 présente les résultats obtenus sur la trace réelle, à comparer avec la figure 2.11. Nous pouvons constater que la modification de la log-vraisemblance, et donc la minimisation du poids apporté par les premières portes des échos, a permis de supprimer le biais. L'estimation bayésienne retrouve alors ces propriétés : pour l'estimation de P_u et τ , le *MLE* approché et l'approche bayésienne fournissent des résultats équivalents. Par contre, pour l'estimation de *swh* nous pouvons observer une diminution notable de la variance des estimations. Cette diminution est mise en évidence avec la courbe « *swh* filtré » (en vert) qui est obtenue à partir d'un filtre médian des estimations successives de *swh* par le *MLE* approché.

D'autres traces réelles ont été testées par la société CLS avec cette nouvelle méthode d'estimations bayésienne confirmant ces résultats.

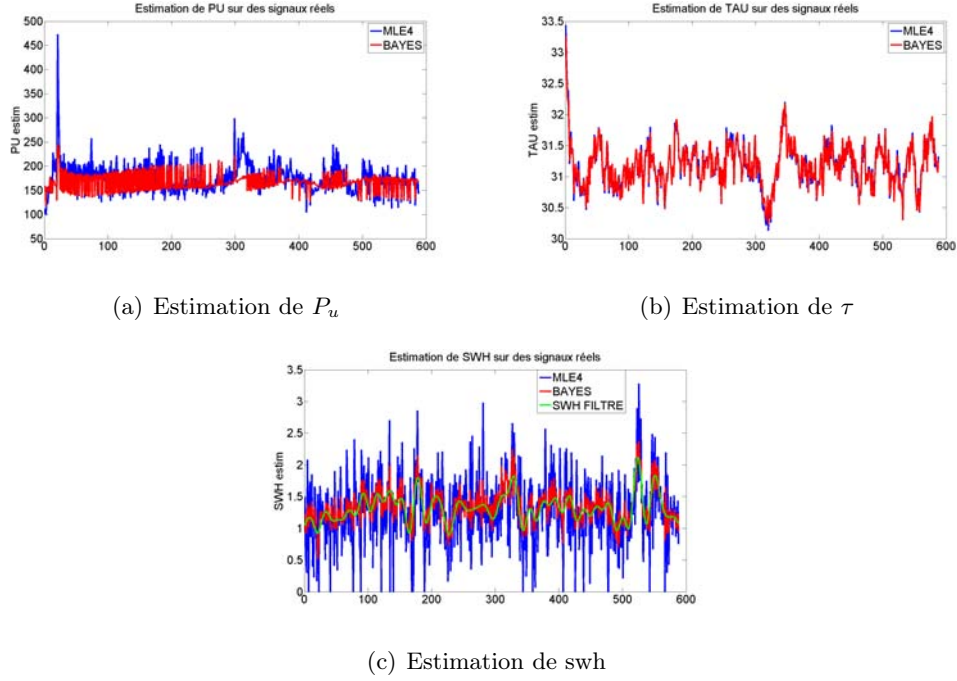


FIG. 2.14 – Estimation de P_u , τ et de swh sur des échos de la trace réelle avec la nouvelle log-vraisemblance.

2.6 Cas particulier du dépointage de l'antenne

2.6.1 Définition

Avec les simulations analysées dans le paragraphe précédent, il a été possible de mettre en avant l'intérêt de l'approche bayésienne. En effet, il est possible d'améliorer sensiblement les estimées pour les paramètres P_u (ou σ_0), τ et de manière significative celles de swh en utilisant l'approche dynamique.

Cependant, dans certains cas d'étude, il est nécessaire d'inclure, dans le modèle de Hayne, le dépointage de l'antenne ξ . Dans ce cas là, le modèle des formes d'onde est plus complexe comme le montre (1.3). Pour rappel, le dépointage de l'antenne symbolise l'angle entre l'orientation du satellite et le nadir (plus courte distance satellite-surface). Cet angle influe directement sur la pente du plateau final d'une forme d'onde mais aussi

sur le front de montée. Plus l'angle est important, plus l'énergie reçue par l'antenne est diffuse. Ce paramètre influe donc significativement sur l'aspect des formes d'onde. La figure 2.15 illustre des formes d'onde de paramètre P_u , τ et swh identiques, mais de ξ différents.

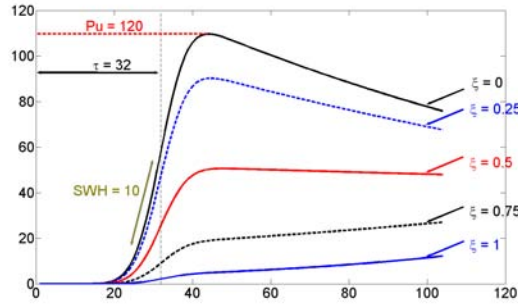


FIG. 2.15 – Influence de ξ sur une forme d'onde.

2.6.2 Estimation à partir du nouveau modèle

Algorithme Tout comme les 3 paramètres principaux, aucune expression analytique simple de la densité de probabilité *a posteriori* de ξ ne peut être déterminée. La méthode d'estimation est la même que pour les autres paramètres. Une chaîne de Markov est construite avec des échantillons distribués selon la fonction de densité de probabilité *a posteriori* de ξ . L'intégration du paramètre ξ dans l'algorithme de Metropolis within Gibbs est illustrée dans l'ALGO. 4.

Loi *a priori* Les lois *a priori* des paramètres P_u , τ et swh sont les lois uniformes définies dans la partie 2.4.1. Observant une symétrie dans le modèle (1.3) pour le paramètre ξ , l'intervalle de définition est donc défini en degré par :

$$\xi \sim \mathcal{U}[0, 1]. \quad (2.20)$$

Simulation Afin d'évaluer la robustesse à la complexification du modèle et aux différentes valeurs du paramètre ξ , les simulations sont effectuées pour des valeurs croissantes

ALGO. 4 Algorithme de « Metropolis within Gibbs à 4 pas de Metropolis-Hastings ».

$$P_u^0 \sim q(P_u)$$

$$\tau^0 \sim q(\tau)$$

$$\text{swh}^0 \sim q(\text{swh})$$

$$\xi^0 \sim q(\xi)$$

POUR $t = 1$ à T FAIRE

$$\phi_{P_u} \sim q(P_u)$$

$$P_u^{t+1} = \begin{cases} \phi_{P_u} \sim \min \left\{ \frac{f(\mathbf{y}|\phi_{P_u}, \tau^t, \text{swh}^t, \xi^t, N_t^t) f(\phi_{P_u})}{f(\mathbf{y}|P_u^t, \tau^t, \text{swh}^t, \xi^t, N_t^t) f(P_u^t)} \frac{q(P_u^t)}{q(\phi_{P_u})}, 1 \right\} \\ P_u^t \text{ sinon} \end{cases}$$

$$\phi_\tau \sim q(\tau)$$

$$\tau^{t+1} = \begin{cases} \phi_\tau \sim \min \left\{ \frac{f(\mathbf{y}|P_u^{t+1}, \phi_\tau, \text{swh}^t, \xi^t, N_t^t) f(\phi_\tau)}{f(\mathbf{y}|P_u^{t+1}, \tau^t, \text{swh}^t, \xi^t, N_t^t) f(\tau^t)} \frac{q(\tau^t)}{q(\phi_\tau)}, 1 \right\} \\ \tau^t \text{ sinon} \end{cases}$$

$$\phi_{\text{swh}} \sim q(\text{swh})$$

$$\text{swh}^{t+1} = \begin{cases} \phi_{\text{swh}} \sim \min \left\{ \frac{f(\mathbf{y}|P_u^{t+1}, \tau^{t+1}, \phi_{\text{swh}}, \xi^t, N_t^t) f(\phi_{\text{swh}})}{f(\mathbf{y}|P_u^{t+1}, \tau^{t+1}, \text{swh}^t, \xi^t, N_t^t) f(\text{swh}^t)} \frac{q(\text{swh}^t)}{q(\phi_{\text{swh}})}, 1 \right\} \\ \text{swh}^t \text{ sinon} \end{cases}$$

$$\phi_\xi \sim q(\xi)$$

$$\xi^{t+1} = \begin{cases} \phi_\xi \sim \min \left\{ \frac{f(\mathbf{y}|P_u^{t+1}, \tau^{t+1}, \text{swh}^{t+1}, \phi_\xi, N_t^t) f(\phi_\xi)}{f(\mathbf{y}|P_u^{t+1}, \tau^{t+1}, \text{swh}^{t+1}, \xi^t, N_t^t) f(\xi^t)} \frac{q(\xi^t)}{q(\phi_\xi)}, 1 \right\} \\ \xi^t \text{ sinon} \end{cases}$$

FIN

de ξ . Les autres paramètres sont générés aléatoirement selon leurs lois *a priori*. L'approche bayésienne et le *MLE* approché sont confrontés en comparant leurs MSEs. Les chaînes de Markov sont obtenues pour un temps de chauffe $Nbi = 1500$ et $Nbr = 6500$ échantillons d'intérêt.

Tout d'abord, en observant les courbes de la figure 2.16 nous constatons que pour des valeurs de $\xi \geq 0.7^\circ$, l'algorithme basé sur *MLE* n'obtient plus d'estimation précise. Cette observation est aussi valable pour l'approche bayésienne mais seulement pour les paramètres τ et swh qui sont directement liés au front de montée. Cela s'explique en observant la figure 2.15, plus ξ est grand et plus il devient difficile de distinguer le front de montée. Il en est de même pour les algorithmes, l'estimation de ces deux paramètres devient trop difficile. Enfin, l'estimation bayésienne des paramètres P_u et ξ est performante.

θ	P_u	τ	swh	ξ	<i>MPSRF</i>
$\sqrt{\widehat{R}}$	1.6418	1.0665	1.0238	1.6505	18.6995

TAB. 2.2 – Potential Scale Reduction Factors.

Convergence Cependant, lorsque l'on se concentre sur les valeurs des critères de convergence du tableau 2.2, pour $\xi = 0.5$, plusieurs remarques sont possibles. Tout d'abord, le critère de convergence paramètre par paramètre indique que les chaînes de Markov des paramètres σ_0 et ξ n'ont pas totalement convergé (critère supérieur à 1.2). De plus, la valeur prise par la version multivariée de ce critère est relativement importante. Or comme le suggère [BG98], le *MPSRF* est plus représentatif de la convergence des chaînes que son homologue univarié (le *PSRF*). Le nombre d'échantillons nécessaires pour obtenir une bonne convergence des chaînes peut donc être une question intéressante. Les figures 2.17 illustrent l'évolution du *MPSRF* en fonction du nombre d'échantillons Nbr utilisés après une temps de chauffe $Nbi = 1500$ itérations. Les figures sont construites à partir de 100 estimations de deux formes d'onde de paramètre P_u, τ et swh identiques, pour $\xi = 0.2$ et $\xi = 0.5$. Pour $\xi = 0.2$ illustré par la figure 2.17(a), en utilisant 20000

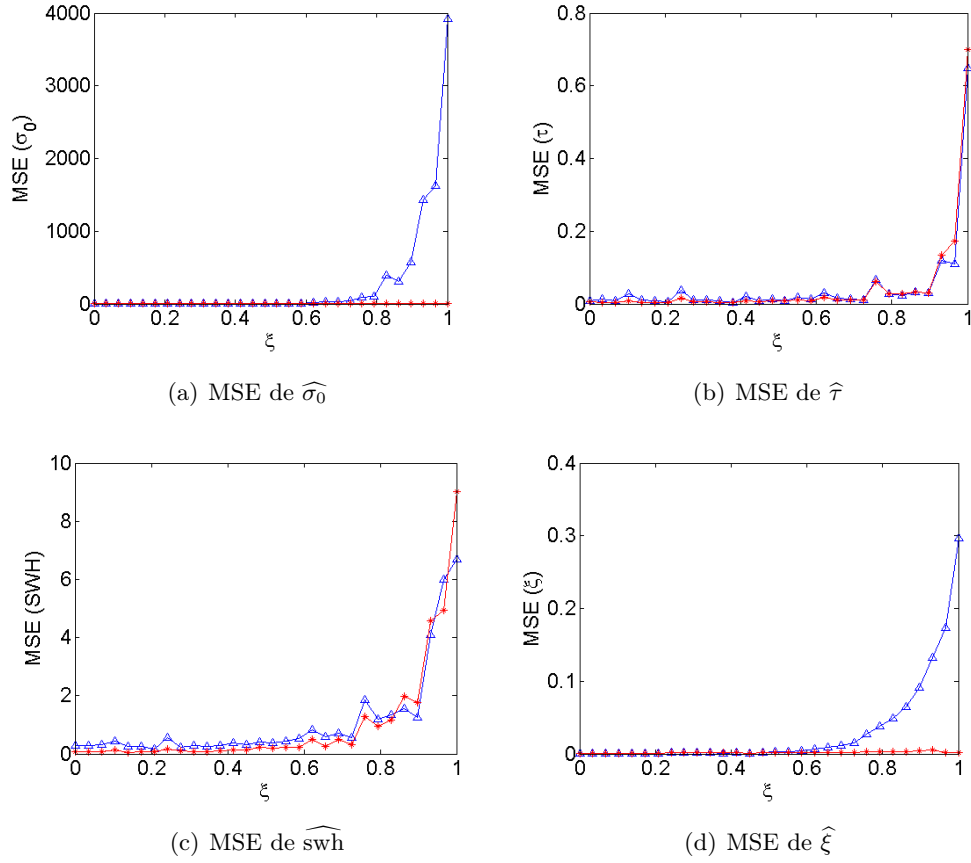


FIG. 2.16 – Comparaison des MSEs en fonction de ξ : $\text{---}\ast\text{---}$ MCMC et $\text{---}\triangle\text{---}$ MLE.

échantillons d'intérêt pour l'estimation, nous pouvons conclure que les chaînes de Markov ont bien convergé. Pour $\xi = 0.5$ comme le montre la figure 2.17(b), en utilisant 90000 échantillons pour l'estimation, le MPSRF se rapproche du maximum de 1.2 requis pour conclure sur la convergence. Le coût calculatoire avec 90000 itérations est alors beaucoup plus important. L'estimation des 3 paramètres principaux de la première partie de l'estimation se fait en une vingtaine de secondes afin de valider la convergence. Pour le cas où le dépointage de l'antenne est pris en compte et $\xi = 0.5$, et afin de valider la convergence des chaînes de Markov (91500 itérations en comptant le temps de chauffe), l'estimation d'une forme d'onde prend alors plus de 7 minutes. L'utilisation d'une loi uniforme

comme loi *a priori* dans le cas du dépointage de l'antenne n'est donc pas une solution envisageable. Cependant, si l'on se place dans une optique de loi *a priori* dynamique (utilisation de lois Gaussiennes), comme cela a été présenté dans la partie précédente, nous pouvons conclure sur la convergence. La dernière figure 2.17(c) représente l'évolution du *MPSRF* en fonction du nombre d'échantillons d'intérêt utilisés dans le cas de loi *a priori* dynamiques. Ce calcul a été effectué à partir de 100 estimations d'une même forme d'onde pour $\xi = 0.5$. La loi *a priori* (ainsi que la loi de proposition) utilisée est une loi normale de moyenne $m = 0.45$ (car on suppose que $\xi = 0.45$ à l'instant $t - 1$) et de variance 0.0025. Comme nous pouvons l'observer, après un « burn-in » de 1500 itérations et une estimation à partir de 6500 échantillons d'intérêt, les chaînes de Markov ont bien convergé vers le même vecteur paramètre $\theta = (P_u, \tau, \text{swh}, \xi)$ car le *MPSRF* est inférieur à 1.2. Dans le cas où le dépointage est estimé, l'approche dynamique est donc mieux adaptée pour garantir la convergence des chaînes et donc les résultats. En plus du fait qu'il est possible de conclure sur la convergence des chaînes de Markov avec l'approche de lois *a priori* dynamiques, la figure 2.18 met en évidence que cette méthode permet de limiter l'erreur de l'estimation bayésienne. Dans cet exemple, l'utilisation de l'approche bayésienne permet une estimation de τ plus précise d'au moins 13cm et de swh d'au moins 38cm. Mais surtout, on constate que pour des valeurs importantes du dépointage (proche de 1), les estimées ne divergent que très peu, contrairement aux estimées basées sur le *MLE*.

2.6.3 Bilan pour le dépointage de l'antenne

Les simulations portant sur le dépointage de l'antenne exposées précédemment ne permettent pas de conclure lorsque les *priors* utilisées sont uniformes. Le principal problème vient du fait que pour différentes valeurs du dépointage, la vitesse de convergence des chaînes de Markov varie. Cette variation de vitesse de convergence influe directement sur le nombre d'échantillons à générer et donc sur le coût calculatoire d'une estimation en

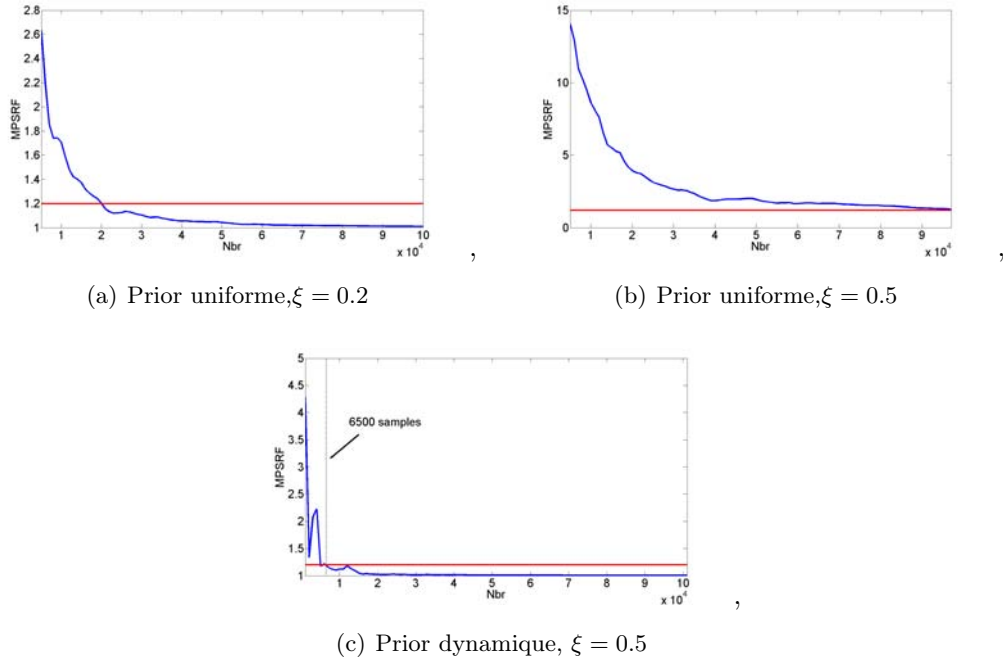
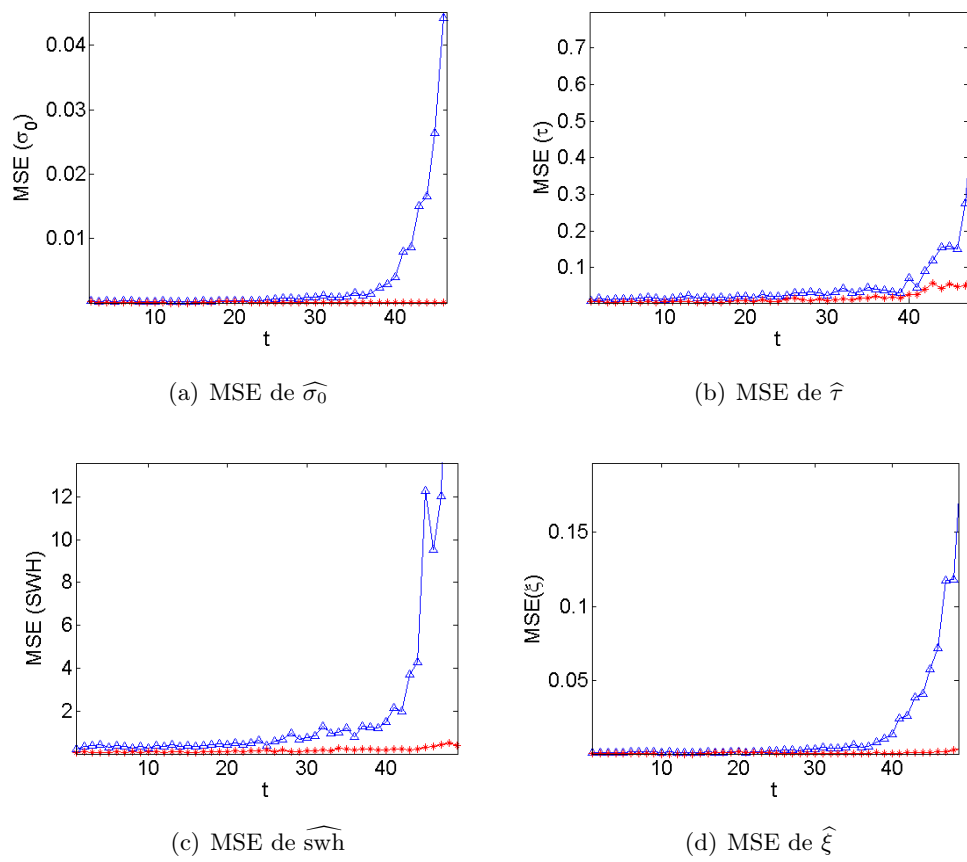


FIG. 2.17 – Critère MPSRF de convergence en fonction du nombre d'échantillons d'intérêt.

utilisant la méthode bayésienne. Même si la convergence n'est pas concluante, l'utilisation de lois *a priori* uniformes permet tout de même d'obtenir de meilleures estimations en comparaison avec celles du *MLE*.

Cependant, comme cela a été fait dans la partie se concentrant sur les 3 paramètres principaux, l'estimation dynamique permet d'utiliser des *priors* beaucoup plus informatives. Ce choix de loi *a priori* permet d'accélérer la vitesse de convergence des chaînes en utilisant des informations issues de la réalité terrain. Ces informations permettent alors à l'approche bayésienne d'obtenir de meilleures performances que le *MLE* surtout dans les cas critiques comme par exemple un dépointage de l'antenne important.

FIG. 2.18 – Comparaison des MSEs en fonction de ξ : $\text{--}\ast\text{--}$ MCMC et $\text{--}\triangle\text{--}$ MLE.

CHAPITRE 3

Classification des formes d'onde altimétriques

Sommaire

3.1	Introduction	59
3.2	Présentation théorique de la classification	61
3.3	Classification de différents types de surface	72
3.4	Problème de classification PISTACH	86

3.1 Introduction

La classification et la reconnaissance de formes sont des disciplines qui ont fait l'objet de nombreux travaux dans la littérature [DH73][DGL96]. Dans cette partie, nous nous intéressons à la classification des formes d'onde altimétriques. Pour rappel, les formes d'onde altimétriques sont des signaux mono-dimensionnels issus d'un altimètre embarqué à bord d'un satellite (ENVISAT/JASON) [CES⁺06]. Plusieurs travaux de recherche ont été menés sur la classification de formes d'onde altimétriques notamment en utilisant des classifieurs sous-optimaux, tels que la distance aux barycentres et la méthode des K plus proches voisins [Sol05], des classifieurs optimaux construits à partir de lois *a priori* décrites par des mélanges de processus de Dirichlet, le classifieur bayésien [DT07][Sol05], ou encore des approches à base de réseaux de neurones [QJR⁺07][Sol05]. Toutes ces méthodes utilisent des paramètres issus de l'altimètre mais aussi des paramètres obtenus à partir d'autres instruments de mesure, comme par exemple le radiomètre. Dans cette partie nous nous intéressons à plusieurs problèmes de classification, en fixant comme

contrainte l'utilisation exclusive de données altimétriques.

Nous nous fixons deux problèmes de classification rencontrés dans la littérature ou suggérés par des rapports internes à la société CLS :

- Nous étudierons dans un premier temps le problème de classification illustré dans [Sol05], à savoir une classification automatique des signaux selon les surfaces étudiées. On définit *a priori* quatre classes : « Déserts », « Eaux et Forêts », « Glaces » et « Océans ». En effet, actuellement, toutes les formes d'onde sont traitées avec un algorithme spécialement adapté aux océans. Cependant, les caractéristiques extraites avec cet algorithme ne trouvent pas systématiquement de sens physique pour les autres surfaces étudiées (Désert, Glaces, ...). Une classification performante de ces types de formes d'onde autoriserait donc d'appliquer un algorithme adapté à chaque type de surface.
- Enfin, nous choisirons une méthode de classification basée sur la forme des échos altimétriques comme cela est présenté dans [Thi08]. Il s'agit de définir des profils types d'échos altimétriques à partir de « squelettes » fréquemment rencontrés par les spécialistes du domaine. Les différentes classes issues de ce traitement ont été définies par la société CLS.

L'objectif de ce chapitre est de présenter les techniques de classification statistique qui semblent adaptées à la reconnaissance des divers types de surface. D'une manière générale, un système de classification, illustré sur la figure 3.1, se compose :

- d'un ensemble de capteurs transmettant l'information en rapport avec les formes d'onde,
- d'une étape de prétraitement dont le rôle est d'extraire des paramètres pertinents pour la classification,
- d'une étape d'apprentissage qui permet de déterminer la règle de décision à mettre en place. Il s'agit d'une étape basée sur l'expertise des signaux,

- d’une règle de décision qui permet la classification automatique des différents signaux altimétriques.

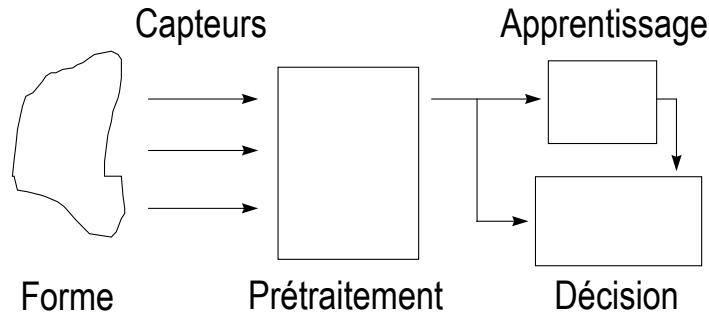


FIG. 3.1 – Schéma de principe pour la classification.

3.2 Présentation théorique de la classification

Nous présentons ici les outils de classification qui seront utilisés sur les différentes applications qui sont détaillées par la suite.

3.2.1 Prétraitement

Dans cette partie, considérons que les signaux altimétriques à classer sont des vecteurs de \mathbb{R}^p , où p représente le nombre de variables. En traitement du signal ou des images, les signaux ou images à classer (les individus) sont souvent constitués d’un nombre important de variables pouvant mener à un problème de surdimensionnement (« curse of dimensionality »). Deux types de prétraitements sont présentés ici qui cherchent à représenter les données de \mathbb{R}^p dans un autre espace de dimension inférieure \mathbb{R}^q , avec $q < p$: l’Analyse en Composantes Principales (ACP) qui fonctionne en mode non supervisé et l’Analyse Factorielle Discriminante (AFD) ou encore appelée Analyse Linéaire Discriminante (ALD ou LDA en anglais), qui utilise un ensemble d’apprentissage constitué de signaux expertisés appartenant à chaque classe (mode supervisé).

Analyse en Composantes Principales des individus

L'Analyse en Composantes Principales (ACP) des individus est une méthode mathématique permettant de transformer un certain nombre p de variables corrélées en un nombre q , $q < p$, de variables décorrélées appelées composantes principales. Il s'agit d'une méthode non supervisée puisqu'elle n'utilise pas l'information apportée par l'expertise des données et ne nécessite pas d'apprentissage.

La détermination du nouvel espace \mathbb{R}^q est tel que la moyenne des carrés des distances entre les individus de \mathbb{R}^p et leurs projections dans \mathbb{R}^q soit minimale. Afin d'éviter les problèmes liés à des variables ayant des dynamiques très différentes, les données de \mathbb{R}^p sont centrées réduites. Soit T la matrice des covariances des individus, il est alors montré [Sap90] que les vecteurs principaux de l'espace \mathbb{R}^q sont les vecteurs propres unitaires associés aux valeurs propres de la matrice T . La variance des projections des individus sur un axe principal est d'autant plus grande que la valeur propre associée à cet axe est grande. Les q variables retenues pour la classification sont alors obtenues par projection sur les q axes associés aux q plus grandes valeurs propres de T appelés axes principaux.

Une fois les vecteurs propres définis, les données sont alors projetées dans le nouvel espace \mathbb{R}^q en utilisant une matrice de transformation construite à partir des vecteurs propres.

Analyse Factorielle Discriminante

L'Analyse Factorielle Discriminante (AFD), est une méthode souvent utilisée en statistique afin de déterminer quelle combinaison linéaire des paramètres permettra la meilleure séparation des classes à analyser. Elle permet aussi la réduction de la taille des données et d'éviter le problème de surdimensionnement. Contrairement à l'ACP, il s'agit d'une méthode supervisée. L'information apportée par l'expertise des signaux d'apprentissage (les labels, indiquant l'appartenance du signal à une classe donnée) est donc nécessaire pour ce type d'analyse.

Cette méthode statistique de séparation des données consiste à minimiser les variances intraclasse et maximiser les variances interclasse des données projetées. Dans le cas de K classes W_i avec $i = 1, \dots, K$, on cherche à maximiser le critère de Fisher défini de la façon suivante :

$$J(u) = \frac{u^t B u}{u^t S u} \quad (3.1)$$

B et S étant deux matrices appelées respectivement matrice de dispersion interclasse et matrice de dispersion intraclasse définies par :

$$B = \sum_{i=1}^K n_i (m_i - m)(m_i - m)^t, \quad (3.2)$$

$$S = \sum_{i=1}^K \sum_{x \in W_i} (x - m_i)(x - m_i)^t, \quad (3.3)$$

n_i étant le nombre d'éléments de W_i , m_i le vecteur moyenne de la classe W_i et m le vecteur moyenne de l'ensemble des signaux de toutes les classes W_i ($i = 1, \dots, K$). Plus $J(u)$ est élevé, plus l'axe u est pertinent. Il a été démontré dans [DH73] que les solutions de ce problème de maximisation sont les $K - 1$ vecteurs propres associées aux $K - 1$ valeurs propres non nulles de la matrice $S^{-1}B$. Plus la valeur propre est importante, plus l'axe u correspondant est discriminant.

3.2.2 Le classifieur bayésien

Le classifieur utilisé est le classifieur du maximum de vraisemblance sous hypothèse gaussienne qui a la particularité de minimiser la probabilité d'erreur. Le vecteur constitué des projections des données sur les axes discriminants sera supposé gaussien. La vraisemblance d'un signal altimétrique pour une classe C_i donnée se calcule à partir de la densité de probabilité f d'une loi gaussienne multivariée de vecteur moyenne M_i de taille q , et de matrice de covariance \sum_i de taille $q \times q$, q représentant le nombre d'axes discriminants utilisés. La règle de décision pour affecter un signal donné à une classe \hat{C} s'écrit alors sous la forme :

$$\hat{C} = \text{Arg} \max_C p(C|\mathbf{x}) = \text{Arg} \max_C \frac{f(\mathbf{x}|C)p(C)}{f(\mathbf{x})}. \quad (3.4)$$

Lorsque les différentes classes sont équiprobables, \hat{C} est la classe qui maximise la vraisemblance des données \mathbf{x} .

3.2.3 La distance de Mahalanobis

Dans le cadre de notre étude, la loi gaussienne sera utilisée comme approximation de $f(\mathbf{x}|C)$. Le classifieur bayésien se réduit alors à des calculs de distances de Mahalanobis suivis d'une règle de distance aux barycentres. La distance de Mahalanobis entre un vecteur \mathbf{x} et une classe composée de plusieurs signaux se définit par [Mah36] :

$$d_{\text{Mahalanobis}}^2(\mathbf{x}, \boldsymbol{\mu}_i) = (\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i), \quad (3.5)$$

$\boldsymbol{\mu}_i$ étant le barycentre de la classe C_i , et Σ_i étant la matrice de covariance de la classe C_i .

3.2.4 Détection d'anomalies

Dans les problèmes de classification présentés par la suite, nous distinguons deux approches :

- soit nous disposons de plusieurs classes que nous traitons simultanément avec le classifieur bayésien,
- soit nous choisissons au préalable d'isoler une classe « saine » importante (contenant ici les échos de Brown) pour ensuite se focaliser sur les classes restantes considérées comme des anomalies vis-à-vis de la classe « saine ». Il s'agit en fait d'une étape de prétraitement permettant d'accélérer la classification. Pour cela nous utiliserons une méthode de détection d'anomalies.

La méthode des machines à vecteurs supports à une classe (appelée « One Class Support Vector Machine » en anglais) introduite par [SPT⁺99] permet notamment la détection d'anomalies. Il s'agit d'une méthode performante et de plus en plus utilisée dans différents domaines tels que le traitement d'images [CZH01], [ZDZ08], la bio-informatique [JZZ07] ou encore la classification de signaux sonores [RD07].

Principe

Contrairement aux algorithmes fonctionnant en mode supervisé, la phase d'apprentissage de la méthode du « One Class SVM » se fait en utilisant seulement les données normales (dites positives). Soit $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ un ensemble d'apprentissage composé de vecteurs à q dimensions appartenant à \mathbb{R}^q . Une interprétation géométrique possible [TD04] de la méthode du « One Class SVM » consiste à trouver la sphère (ou l'hyper-sphère) de volume minimum contenant l'ensemble \mathcal{X} des signaux d'apprentissage, il s'agit du « Support Vector Data Description » (SVDD). Pour cela, nous caractérisons la sphère par son centre \mathbf{a} et son rayon R , puis cherchons à minimiser R^2 . Pour tout $\mathbf{x}_i \in \mathcal{X}$, la fonction de minimisation se définit alors :

$$F(R, \mathbf{a}) = R^2, \quad (3.6)$$

avec comme contrainte :

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2. \quad (3.7)$$

Afin de permettre l'existence « d'outliers » dont la distance au point \mathbf{a} n'est pas strictement inférieure à R^2 , des variables de relaxation ξ_i (en anglais « slack variables ») sont introduites. Notre fonction de minimisation devient donc :

$$F(R, \mathbf{a}) = R^2 + C \sum_i \xi_i, \quad (3.8)$$

où C est une constante permettant de régler l'importance de la pénalisation et où les contraintes suivantes doivent être respectées :

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i \quad (3.9)$$

$$\xi_i \geq 0, \forall i. \quad (3.10)$$

Le problème de minimisation se résume ainsi :

$$\min_R \left[R^2 + C \sum_i \xi_i \right], \text{ avec } \|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i, \forall i. \quad (3.11)$$

Schölkopf [SPT⁺99] propose une autre approche géométrique pour la détection d'anomalies. Il s'agit de déterminer un hyperplan défini par un vecteur normal $\boldsymbol{\omega}$ et sa distance à l'origine ρ , afin de séparer l'ensemble d'apprentissage \mathcal{X} de l'origine avec une marge maximum. Les « outliers » sont toujours gérés avec les variables de relaxation ξ_i , mais Schölkopf introduit un paramètre ν quantifiant le nombre d'« outliers » autorisés dans le problème de minimisation en fonction du nombre N de signaux d'apprentissage. Le problème d'optimisation proposé se définit par :

$$\min_{\boldsymbol{\omega}, \rho} \left[\frac{1}{2} \|\boldsymbol{\omega}\|^2 - \rho + \frac{1}{N\nu} \sum_i \xi_i \right], \quad (3.12)$$

avec pour contraintes :

$$\boldsymbol{\omega}^T \mathbf{x}_i \geq \rho - \xi_i, \quad \forall i \quad \xi_i \geq 0. \quad (3.13)$$

Avec la condition $\|\boldsymbol{\omega}\| = 1$, une formulation équivalente de ce problème peut s'écrire :

$$\max_{\rho} \left[\rho - \frac{1}{N\nu} \sum_i \xi_i \right], \quad (3.14)$$

avec les contraintes :

$$\boldsymbol{\omega}^T \mathbf{x}_i \geq \rho - \xi_i, \quad (3.15)$$

$$\xi_i \geq 0, \quad \forall i, \quad (3.16)$$

$$\|\boldsymbol{\omega}\| = 1. \quad (3.17)$$

Or il est expliqué dans [TD04] que lorsque toutes les données liées à la formulation du « Support Vector Data Description » sont normalisées, le problème peut se formuler de la manière suivante :

$$\min_{R'} \left[R'^2 + C' \sum_i \xi_i'^2 \right], \quad (3.18)$$

avec les contraintes :

$$\|\mathbf{x}_i' - \mathbf{a}'\|^2 \leq R'^2 + \xi_i', \quad \forall i, \quad (3.19)$$

où \mathbf{x}_i' et \mathbf{a}' sont des vecteurs normalisés. En reformulant alors le problème d'optimisation, nous obtenons :

$$\max_{R'} \left[-R'^2 + C' \sum_i \xi_i' \right], \quad (3.20)$$

avec comme nouvelle contrainte :

$$2 [(\mathbf{a}')^T \cdot \mathbf{x}_i'] \geq 2 - R'^2 - \xi_i', \forall i. \quad (3.21)$$

En effectuant un simple changement de variables :

$$\boldsymbol{\omega} = 2\mathbf{a}', \quad (3.22)$$

$$\rho = 2 - R'^2, \quad (3.23)$$

$$\frac{1}{N\nu} = C', \quad (3.24)$$

$$\xi_i = \xi_i', \quad (3.25)$$

le problème d'optimisation se reformule de la manière suivante :

$$\max_{\rho} \left[-2 + \rho - \frac{1}{N\nu} \sum_i \xi_i \right], \quad (3.26)$$

sous les contraintes :

$$\boldsymbol{\omega}^T \mathbf{x}_i \geq \rho - \xi_i, \quad (3.27)$$

$$\xi_i \geq 0, \forall i, \quad (3.28)$$

$$\|\boldsymbol{\omega}\| = 2. \quad (3.29)$$

Les deux problèmes d'optimisation ont donc des solutions similaires comme l'illustre la figure 3.2. Nous utiliserons donc par la suite la formulation de [SPT⁺99]. Le problème de cette approche réside dans l'hypothèse de données séparables par un hyperplan. Lorsque cela n'est pas le cas, l'article [SPT⁺99] propose de transformer les données à l'aide d'une fonction non linéaire ϕ dans le but de les projeter dans un espace où elles seront linéairement séparables. Une fonction noyau κ est alors utilisée, comme le noyau gaussien, tel que :

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle, \quad (3.30)$$

$$\kappa(\mathbf{x}, \mathbf{y}) = \exp \left(\frac{-\|\mathbf{x} - \mathbf{y}\|^2}{c} \right), \quad (3.31)$$

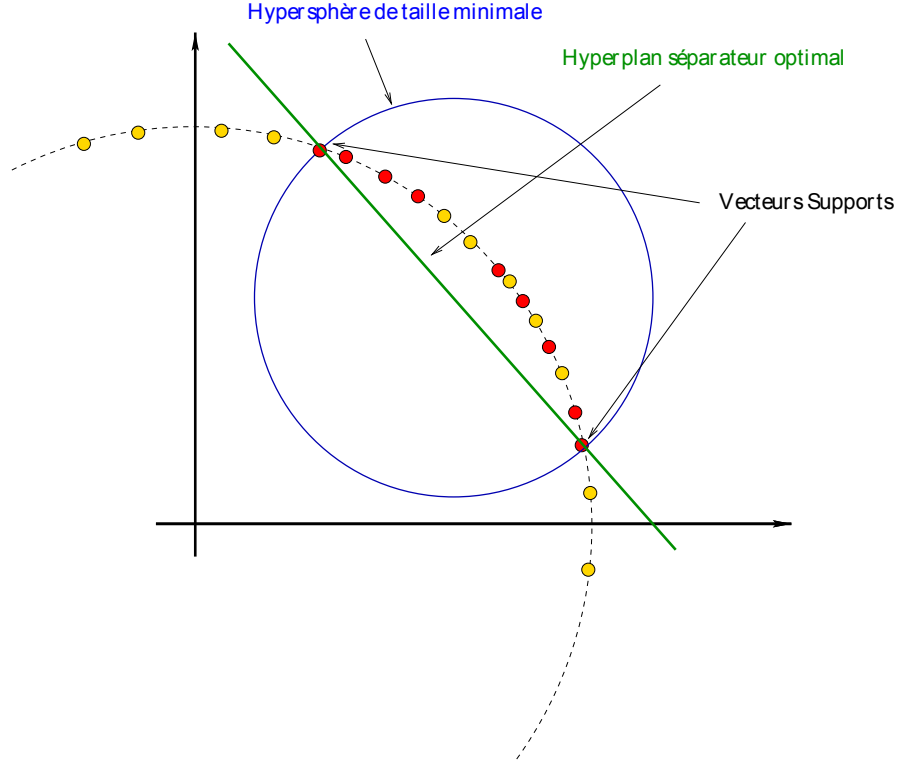


FIG. 3.2 – Schéma de principe pour la détection d'anomalies.

c étant défini comme la largeur du noyau. En écrivant le problème de détection d'anomalies comme un problème d'optimisation sous contraintes [SPT⁺99] donne la fonction de décision D qui détermine si un objet \mathbf{z} (un vecteur à q dimensions) est une anomalie ou pas. La fonction D est définie par :

$$D(\mathbf{z}) = \text{sign} \left(\sum_i \alpha_i \kappa(\mathbf{x}_i, \mathbf{z}) - \rho \right), \quad (3.32)$$

où ici, les vecteurs \mathbf{x}_i sont les vecteurs supports, c'est à dire les signaux d'apprentissage pour lesquels les coefficients α_i sont non nuls. Cette fonction de décision correspond en fait à la somme des distances entre les vecteurs supports \mathbf{x}_i et le vecteur \mathbf{z} , pondérées par les coefficients α_i . Le paramètre ρ ramène cette fonction à un calcul de distance entre \mathbf{z} et l'origine.

Le critère d'alignement

L'utilisation de la méthode SVM nécessite au préalable de choisir la famille de noyau utilisée pour κ (gaussienne, polynômiale, exponentielle, ...). Pour déterminer le(s) paramètre(s) du noyau, l'approche la plus directe est de se baser sur des validations croisées. Il existe également des méthodes de sélection de noyau [Pot07] (mesure résiduelle, alignement, mesure de séparabilité des classes, ...). Dans le cadre de la méthode « One Class SVM », le critère d'alignement semble l'approche la plus adaptée, mais impose un apprentissage en mode supervisé car il est nécessaire de disposer de données étiquetées comme « anomalies ».

En effet, l'utilisation du critère d'alignement demande au préalable de définir un noyau idéal. [NCK02] utilise les libellés des signaux d'apprentissage \mathbf{y} regroupés dans le vecteur \mathbf{z} afin de définir ce noyau idéal :

$$\begin{aligned}\kappa^*(\mathbf{y}_i, \mathbf{y}_j) &= 1 \quad \text{si } \mathbf{y}_i \text{ et } \mathbf{y}_j \text{ appartiennent à la même classe,} \\ &= -1 \quad \text{sinon,}\end{aligned}$$

c'est à dire :

$$\kappa^* = \mathbf{z}\mathbf{z}'. \quad (3.33)$$

Afin de mesurer l'alignement entre ce noyau idéal κ^* et un noyau quelconque κ de même famille, nous définissons respectivement K^* et K , les matrices de Gram respectivement de κ^* et κ , des matrices semi-définies positives dont les termes généraux peuvent s'écrire $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ et $K_{ij}^* = \kappa^*(\mathbf{x}_i, \mathbf{x}_j)$. L'alignement entre les deux noyaux à partir d'un ensemble d'apprentissage \mathcal{X} se définit alors par [NCK02] :

$$\hat{A}(\mathcal{X}, \kappa^*, \kappa) = \frac{\langle K^*, K \rangle_F}{\sqrt{\langle K^*, K^* \rangle_F \langle K, K \rangle_F}} \quad (3.34)$$

$\langle \mathbf{X}, \mathbf{Y} \rangle_F = \sum_i \sum_j X_{ij} Y_{ij}$ représente le produit scalaire de Frobenius entre les matrices \mathbf{X} et \mathbf{Y} .

Un exemple

Considérons des signaux bidimensionnels \mathbf{y} caractérisés par les variables aléatoires X_1 et X_2 . La figure 3.3 permet de visualiser la répartition de nos signaux de la classe normale (en rouge) et des signaux de la classe anormale (en bleu). Il apparaît clairement qu'une simple détection d'anomalies en utilisant un seuil mènera à une séparation peu performante des données.

S'agissant d'un exemple, nous utiliserons l'ensemble de nos signaux afin de construire la méthode SVM. Nous définissons avant tout la famille de noyaux de la méthode SVM comme étant la famille des noyaux gaussiens et nous fixons le taux d'« outliers » $\nu = 10\%$. Pour déterminer l'unique paramètre de ce noyau, sa largeur c (appelée parfois variance), nous allons calculer la valeur du critère d'alignement pour plusieurs valeurs de c et observer son évolution. La figure 3.4 nous permet d'observer cette évolution.

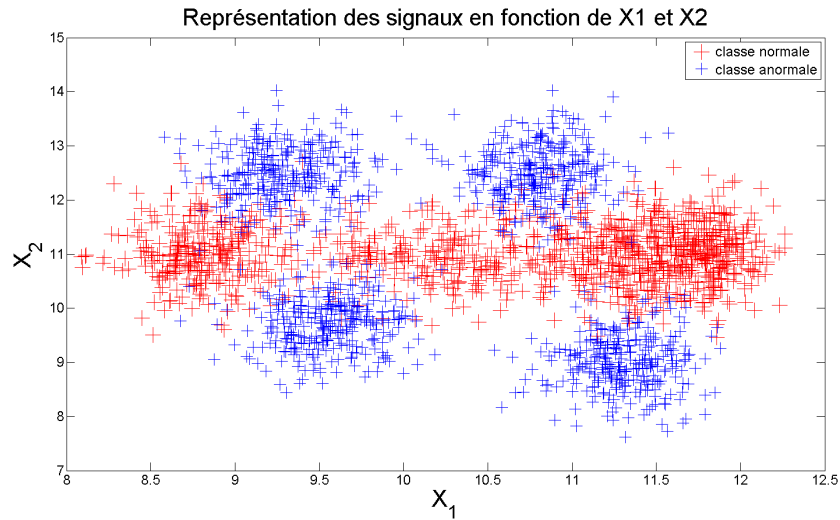


FIG. 3.3 – Représentation des signaux en fonction de X_1 et X_2 .

Nous constatons alors qu'il existe un maximum à cette fonction. La figure 3.5 affiche le résultat obtenu à l'aide de la méthode SVM associée au critère d'alignement minimum. On constate que plusieurs micro-zones de séparations sont déterminées, morcelant beaucoup

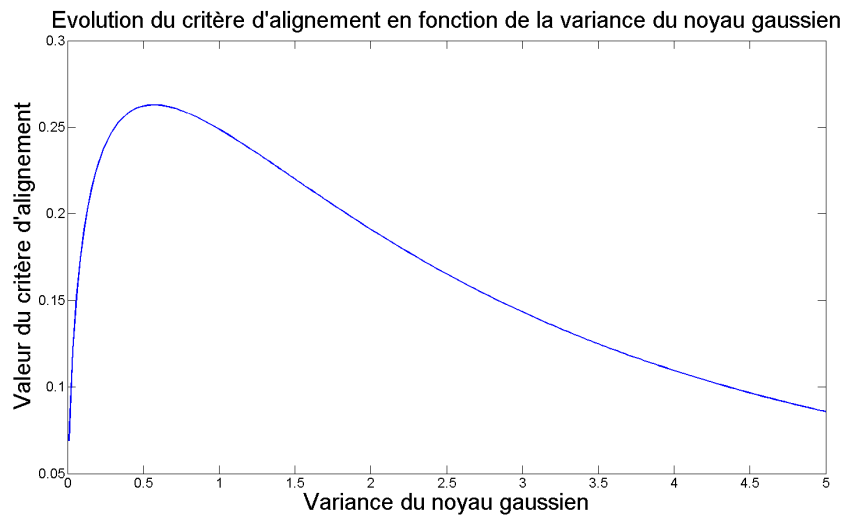


FIG. 3.4 – Illustration de l'évolution du critère d'alignement en fonction de la variance du noyau gaussien du SVM.

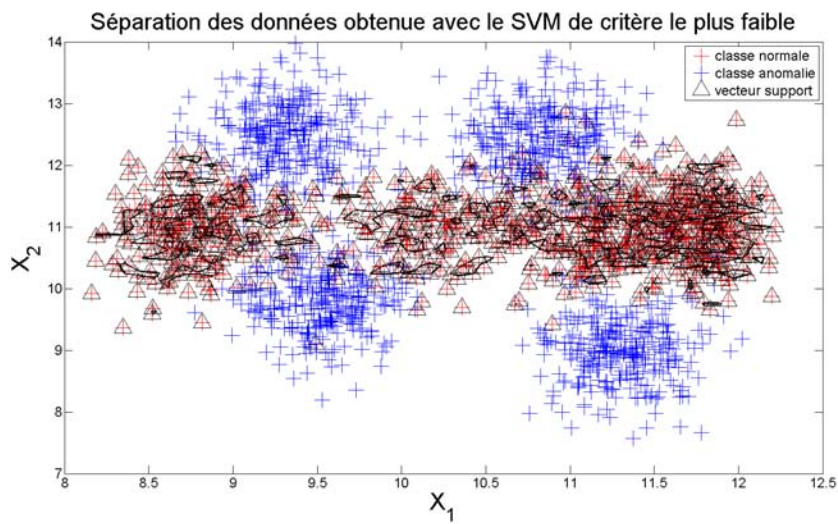


FIG. 3.5 – Résultats de l'utilisation des SVMs pour le critère d'alignement le plus faible.

trop les données de la classe normale. Cependant, lorsque nous utilisons la largeur de noyau c correspondant au critère d'alignement maximum, la figure 3.6 montre que la

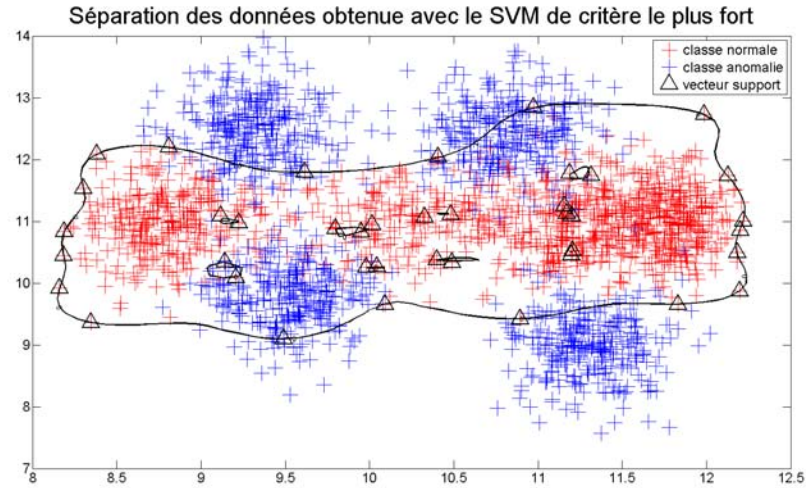


FIG. 3.6 – Résultats de l'utilisation des SVMs pour le critère d'alignement le plus fort.

séparation des données est plus homogène laissant supposer de meilleures performances en terme de classification.

3.3 Classification de différents types de surface

3.3.1 Présentation du problème

Afin de traiter le premier problème de classification, la société CLS nous a fourni une base de données de signaux expertisés issus du satellite ENVISAT. La figure 3.7 illustre la diversité des formes d'onde rencontrées selon les types de surface considérés. Il est à noter cependant que les échos obtenus après réflexion sur des océans ont une forme relativement homogène. En complément des formes d'onde, nous disposons de paramètres liés à la forme des échos altimétriques obtenus par une méthode d'estimation récursive par maximum de vraisemblance (appelée « retracking » [CES⁺06]). Certains paramètres sont calculés écho par écho, avec une fréquence de 18Hz, tandis que d'autres

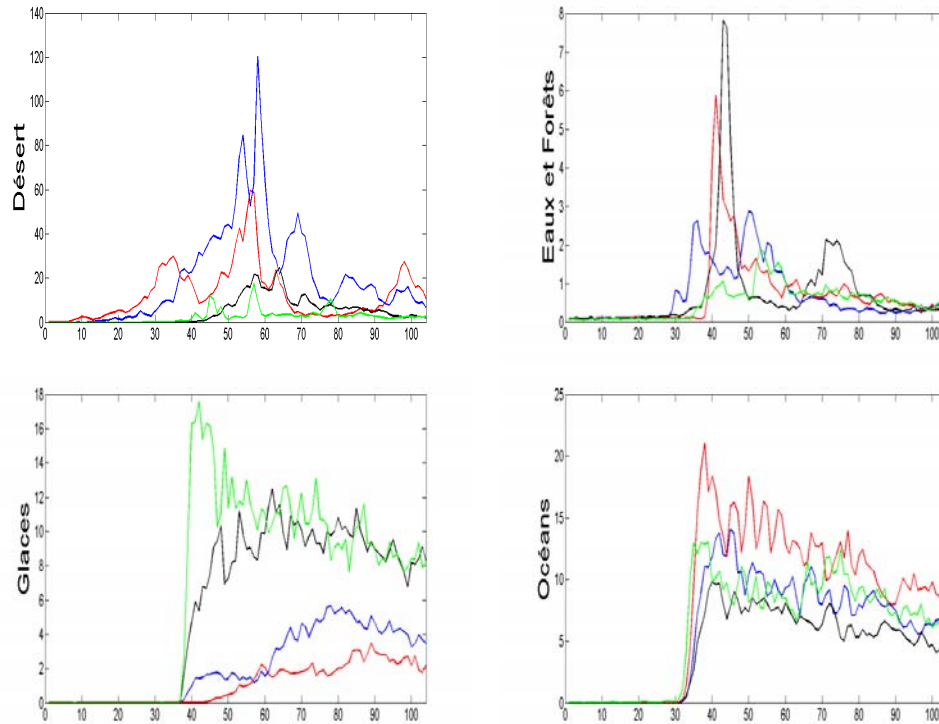


FIG. 3.7 – Exemples de formes d’onde altimétriques selon leurs types de surface.

sont déterminés à partir d’une moyenne de 20 formes d’onde, avec une fréquence de 1Hz.

Voici le descriptif de ceux sélectionnés pour le traitement :

- `ku_range_1Hz`, est une estimation de la distance satellite-surface corrigée,
- `ku_swh_1Hz`, hauteur moyenne des vagues,
- `ku_sigma0_1Hz`, coefficient de rétrodiffusion,
- `elevation_1Hz`, est la différence entre le « range » et l’altitude du satellite. Ce paramètre n’est soumis à aucune correction et permet une première estimation de l’altitude de la surface analysée,
- `ku_peakiness_1Hz`, est une mesure de présence de pic. Cette détection se fait avec le rapport entre le maximum de l’écho et sa moyenne,
- `ku_range_18Hz`,

- elevation_18Hz.

À ces paramètres, nous avons rajouté pour chaque écho altimétrique X composé de $n = 104$ points (appelés portes) les paramètres suivants [Thi08] :

$$\begin{aligned} \text{peakiness_18Hz} &= \frac{\text{Max}(X_i)}{\bar{X}}, \\ \text{kurtosis_18Hz} &= \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}, \\ \text{skewness_18Hz} &= \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\hat{\sigma}} \right)^3, \end{aligned}$$

où \bar{X} et $\hat{\sigma}^2$ sont respectivement la moyenne et la variance estimées de l'écho X .

3.3.2 Traitement des données

De par le nombre et la diversité des données, la classification peut s'appliquer sur différentes combinaisons des paramètres mais aussi selon l'étape de prétraitement choisie. À titre d'exemple, nous donnons ici quelques possibilités comme vecteur en entrée du classifieur :

- les formes d'onde (104 échantillons) acquises tous les 1/18s, chaque porte étant considérée comme un paramètre,
- les formes d'onde moyennées par paquet de 20 échos,
- une combinaison de paramètres caractérisant les formes d'onde comme proposé dans le paragraphe précédent (calculés écho par écho ou par paquets de 20 échos),
- la projection par ACP ou AFD d'un des 3 types de vecteur ci-dessus sur un nombre d'axes discriminants à choisir.

Toutes ces possibilités ont été testées. Dans ce qui suit, nous présentons uniquement les choix qui ont conduit aux meilleurs résultats de classification.

3.3.3 Résultats pour des formes d'onde à 1Hz

Tout d'abord nous nous intéressons à la classification des formes d'onde altimétriques par paquets de 20 échos, en utilisant les 5 paramètres décrits précédemment concernant l'extension 1Hz (ku_range_1Hz , ku_swh_1Hz , ku_sigma0_1Hz , $elevation_1Hz$, $ku_peakiness_1Hz$). Nous procédons à une ACP des individus. Les projections des formes d'onde sur les deux premiers axes principaux sont représentées sur la figure 3.8. Comme nous pouvions nous y attendre en observant la figure 3.7, les projections des échos issus des océans sont très regroupées à la différence des autres types de surface. En opposition, les projections de la classe « Glaces » sont dispersées dans les domaines voisins.

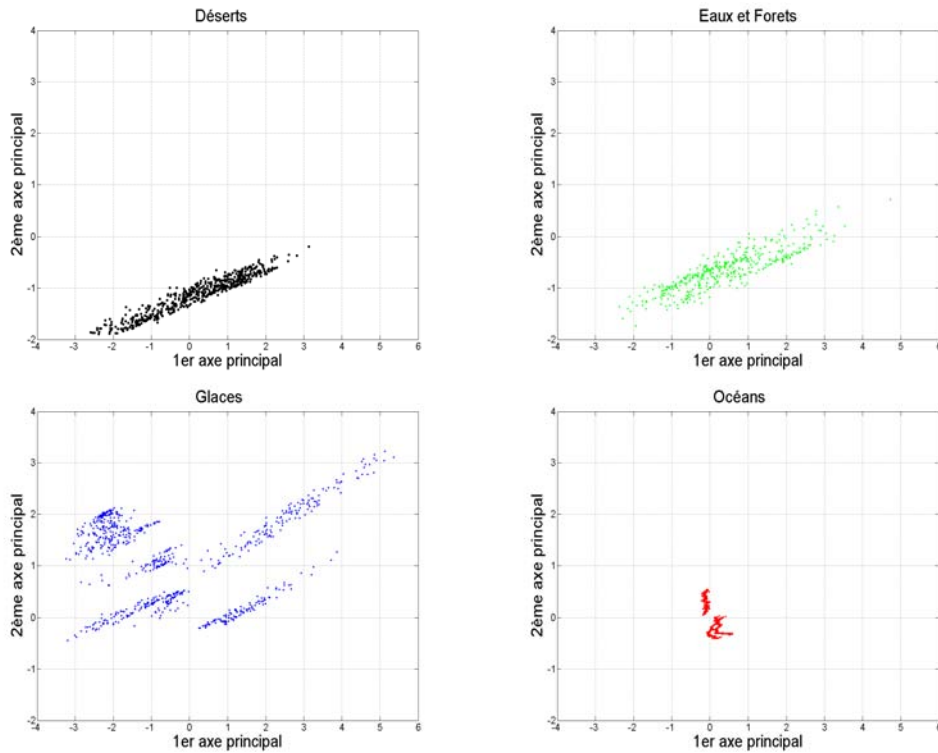


FIG. 3.8 – ACP des formes d'onde altimétriques pour des paramètres déterminés à une fréquence de 1Hz.

leurs pouvoirs discriminants sont mesurés à l'aide des valeurs propres de la matrice de covariance des échos. La figure 3.9 nous permet de constater qu'une représentation sur seulement 2 axes permet d'utiliser approximativement 65% de l'information utile. Dans

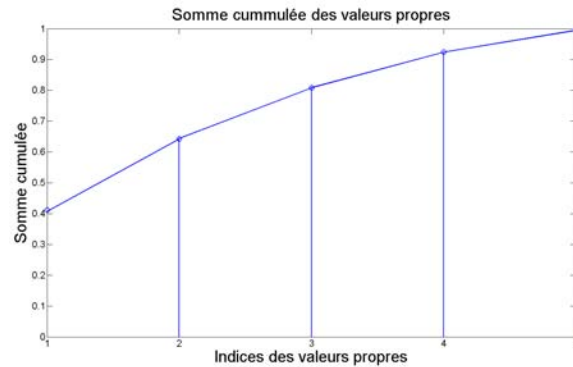


FIG. 3.9 – Somme cumulée des valeurs propres de l'ACP pour des paramètres obtenus à une fréquence de 1Hz.

ce cas précis, nous n'utilisons pas l'ACP dans le but de réduire la dimension du vecteur (qui n'est pas grande), mais plutôt pour décorréler les données. Afin d'utiliser 100% de l'information, nous nous concentrons donc sur une classification basée sur la totalité des 5 axes principaux. Utilisant la moitié de nos signaux (choisis aléatoirement) pour obtenir la matrice de transformation des données issues de l'ACP, l'autre moitié est utilisée pour effectuer une classification des échos. Le tableau 3.1 illustre la matrice de confusion obtenue. Les valeurs de la diagonale de cette matrice de confusion sont très bonnes traduisant une méthode performante de classification des échos par paquets de 20 formes d'onde. De la même manière nous effectuons une AFD sur les paramètres évalués à la fréquence de 1Hz. Les projections des formes d'onde sur les deux axes les plus discriminants sont représentées sur la figure 3.11. Malgré le fait que les axes discriminants soient des combinaisons linéaires des paramètres, il est intéressant d'observer l'information apportée par chacun des paramètres : nous pouvons constater sur la figure 3.10(a) que les paramètres contenant la majorité de l'information nécessaire à la classification sont `ku_range_1Hz`

		affectations			
origines		Déserts	Eaux et Forêts	Glaces	Océans
	Déserts	98.17	1.83	0	0
	Eaux et Forêts	0.48	98.07	1.45	0
	Glaces	0	0	100	0
	Océans	0	2.07	0	97.93

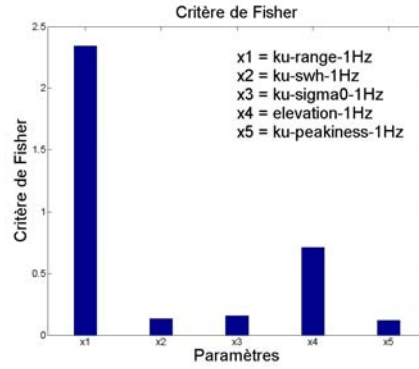
TAB. 3.1 – Table de confusion du classifieur bayésien en % après l’utilisation de l’ACP sur les paquets de 20 échos (calculés à la fréquence de 1Hz).

et elevation_1Hz. La concentration des individus de la classe « Océans » est toujours observable ainsi que l’importance de la variance intraclasse de la classe « Glaces ». Ces deux axes utilisés pour la représentation totalisent environ 93% de l’information utile (figure 3.10(b)) et la classification s’effectue en utilisant les 3 axes de l’AFD. Même si dans grand nombre de cas l’AFD permet une meilleure séparation des données qu’une ACP, le tableau de confusion 3.2 met en avant un dysfonctionnement de la méthode.

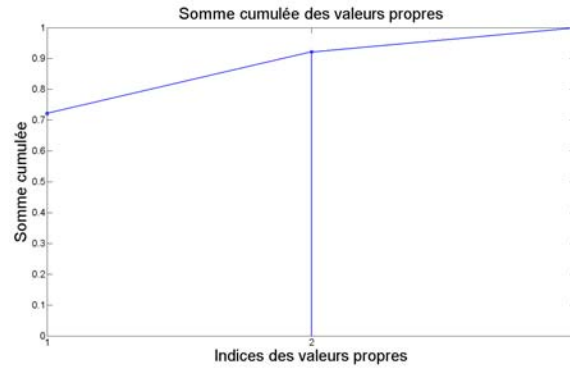
		affectations			
origines		Déserts	Eaux et Forêts	Glaces	Océans
	Déserts	95.25	4.75	0	0
	Eaux et Forêts	14.49	84.54	0.97	0
	Glaces	0	0	100	0
	Océans	0	0.83	0.83	98.34

TAB. 3.2 – Table de confusion du classifieur en % après l’utilisation d’une AFD sur les paramètres des échos calculés en 1Hz.

En effet, une AFD détermine les axes permettant de minimiser les variances intraclasse (afin d’obtenir des classes concentrées) mais aussi de maximiser la variance interclasse (afin de séparer au mieux les classes les unes des autres). Or comme nous pouvons l’observer sur la figure 3.11, la classe des « Glaces » se définit par la variance intraclasse la plus importante mais aussi par la distinction de deux groupes d’individus. Cette grande



(a) Information de Fisher des paramètres.



(b) Somme cumulée des valeurs propres.

FIG. 3.10 – Information de Fisher et somme cumulée des valeurs propres issues de l'AFD pour des paramètres obtenus à une fréquence de 1Hz.

diversité au sein de la classe pénalise les optimisations de l'AFD et principalement celle des variances interclasses.

Après une analyse plus fine des données altimétriques, nous créons donc à partir de la classe « Glaces » deux nouvelles classes : « Glaces Nord » et « Glaces Sud ». Cet éclatement de la classe « Glaces » permet une meilleure séparation des données en utilisant une AFD grâce à une meilleure homogénéité des classes illustrée sur la figure 3.12. Ainsi, les performances du classifieur sont revues à la hausse comme l'illustre le tableau de confusion 3.3.

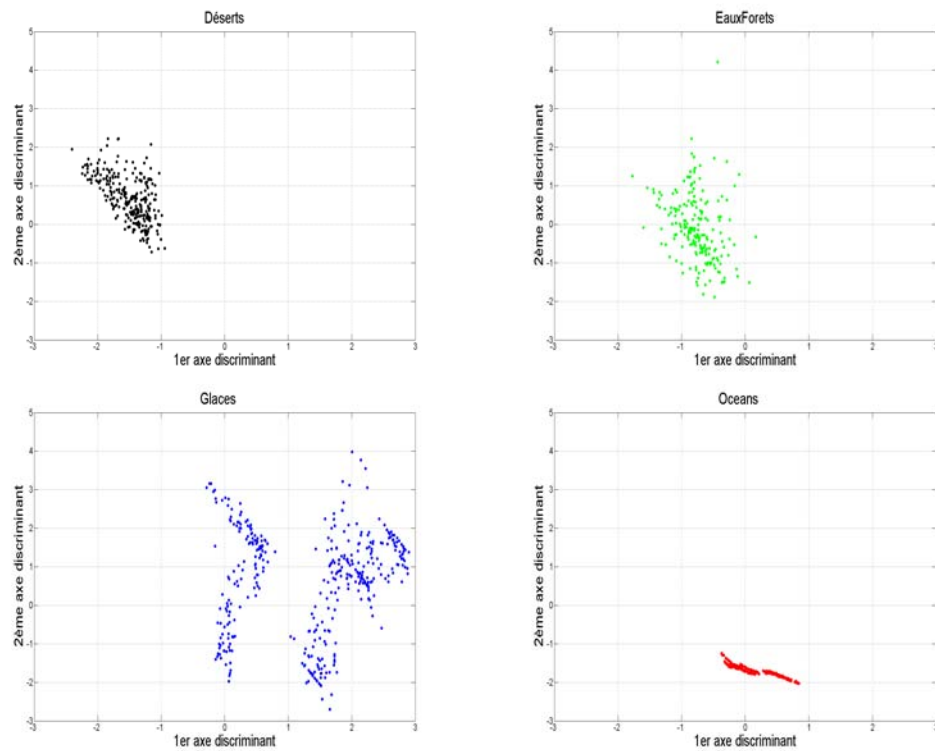


FIG. 3.11 – AFD des formes d’onde altimétriques pour des paramètres déterminés à une fréquence de 1Hz.

affectations						
origines		Déserts	Eaux et Forêts	Glaces Nord	Glaces Sud	Océans
	Déserts	98.36	1.64	0	0	0
	Eaux et Forêts	2.17	96.86	0.97	0	0
	Glaces Nord	0	0	100	0	0
	Glaces Sud	0	0	0	100	0
	Océans	0	1.66	0	0	98.34

TAB. 3.3 – Table de confusion du classifieur en % après l’utilisation d’une AFD sur les paramètres des échos calculés en 1Hz suite à la composition de 2 classes de Glaces.

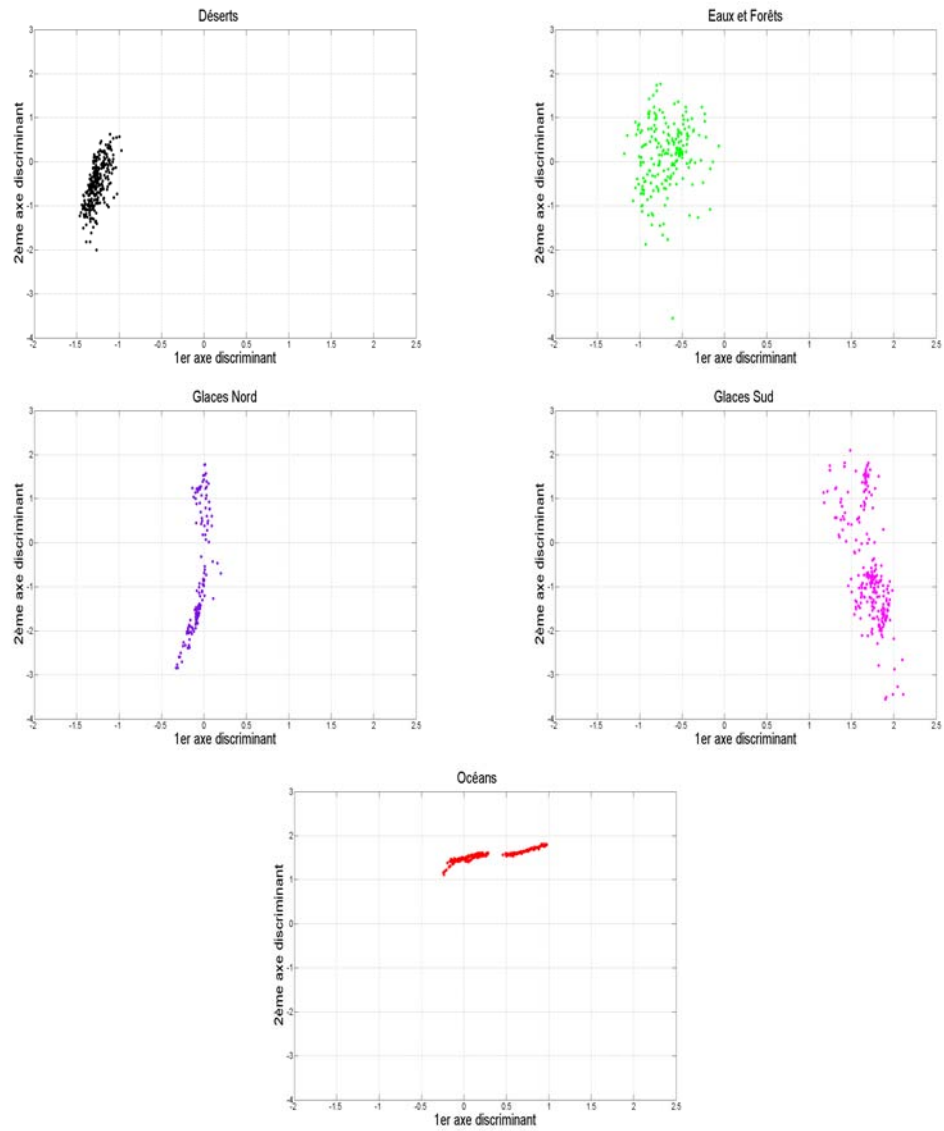


FIG. 3.12 – AFD des formes d'onde altimétriques pour des paramètres déterminés à une fréquence de 1Hz dans le cas de 2 classes de Glaces.

3.3.4 Résultats pour des formes d'onde à 18Hz

Nous nous intéressons ensuite à la classification des formes d'onde altimétriques écho par écho. Les paramètres utilisés sont ceux avec l'extension 18Hz (`ku_range_18Hz`, `elevation_18Hz`) fournis par la société CLS mais aussi les paramètres que nous avons calculés pour chaque écho (`peakiness_18Hz`, `kurtosis_18Hz`, `skewness_18Hz`) ce qui représente un total de 5 paramètres. Nous commençons par appliquer une ACP des individus dont les projections sur les deux axes les plus discriminants sont illustrées sur la figure 3.13. La figure 3.14 nous permet de dire que ces deux axes représentent 85% de l'information disponible. De la même manière que pour le traitement par paquet de 20 échos, la classe « Océan » est très compacte tandis que la classe « Glaces » se caractérise par 4 lots d'individus. Afin d'utiliser la totalité des données accessibles, la classification se fait sur les 5 axes discriminants. Le tableau de confusion 3.4 met en évidence une baisse des performances lorsqu'il s'agit de distinguer les échos issus de la classe « Déserts » des échos de la classe « Eaux et Forêts ».

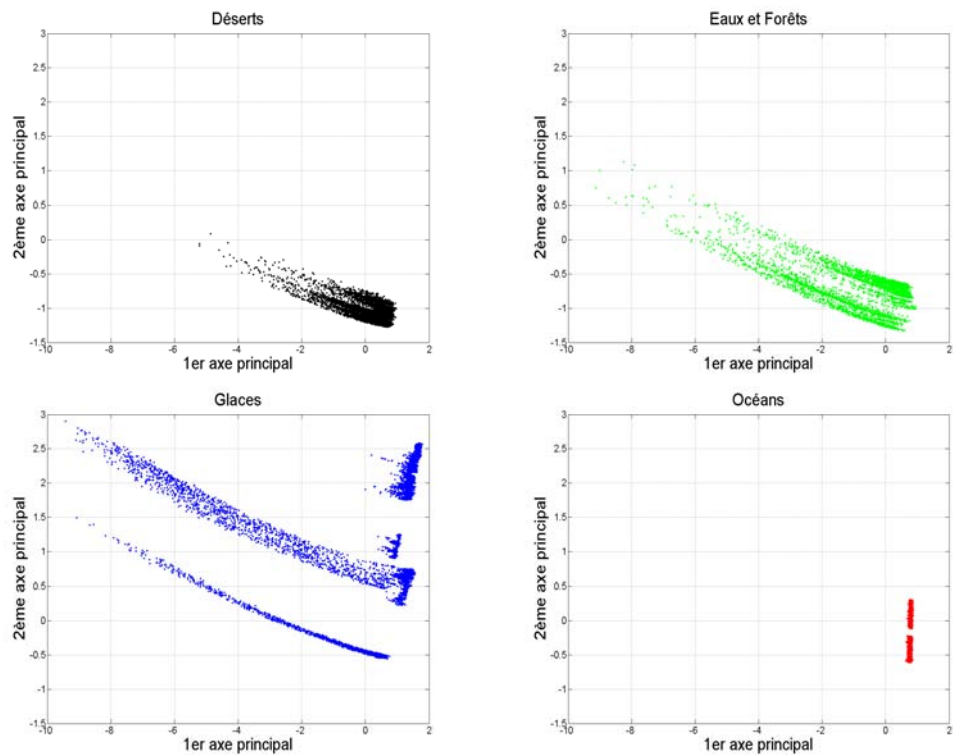


FIG. 3.13 – ACP des formes d'onde altimétriques pour des paramètres déterminés en 18Hz.

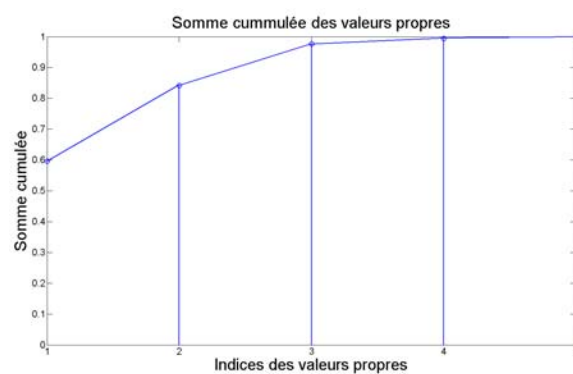
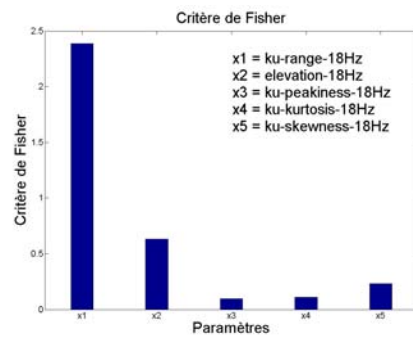


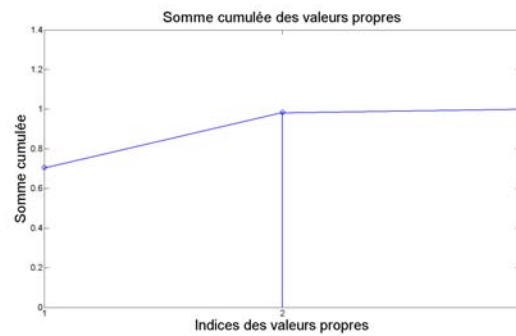
FIG. 3.14 – Somme cumulée des valeurs propres de l'ACP pour des paramètres obtenus à une fréquence de 18Hz.

		affectations			
origines		Déserts	Eaux et Forêts	Glaces	Océans
	Déserts	95.96	3.55	0.45	0.04
	Eaux et Forêts	16.11	81.72	2.17	0
	Glaces	0	9.99	99.39	0.62
	Océans	0	0.04	0.02	99.94

TAB. 3.4 – Table de confusion du classifieur en % après l'utilisation de l'ACP écho par écho.



(a) Information de Fisher des paramètres.



(b) Somme cumulée des valeurs propres.

FIG. 3.15 – Information de Fisher et somme cumulée des valeurs propres issues de l'AFD.

Nous nous intéressons alors aux résultats de la classification obtenus après une AFD. Nous constatons en observant le critère de Fisher des paramètres sur la figure 3.15(a), calculé au cours de l'AFD, que l'information est principalement contenue dans les paramètres `ku_range_18Hz` et `elevation_18Hz`. La figure 3.16 représente les projections des formes d'onde sur les deux axes les plus discriminants après l'application d'une AFD. La forte concentration des échos océaniques est toujours observable ainsi que la grande dispersion des glaces. Cependant, nous pouvons constater que cette représentation sur deux axes ne sera pas suffisante pour la classification même si ces axes correspondent à 98% de l'information comme illustré sur la figure 3.15(b). En effet, avec cette représentation des données, les classes « Glaces », « Eaux et Forêts » et « Déserts » engendrent des projections qui se recouvrent. Nous décidons alors d'utiliser les 3 axes discriminants

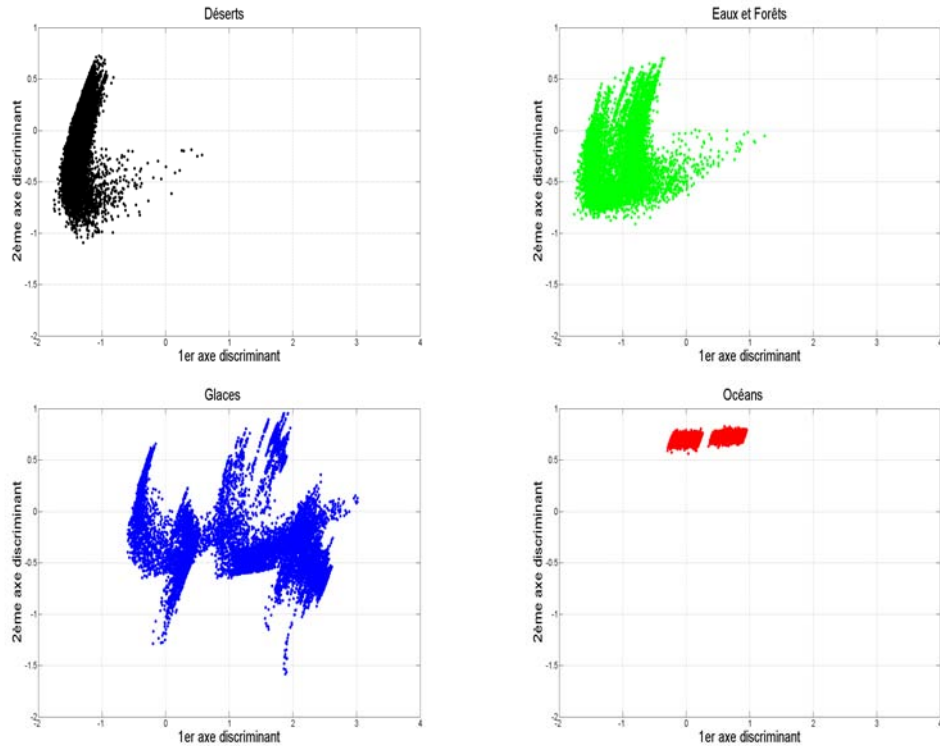


FIG. 3.16 – AFD des formes d'onde altimétriques.

		affectations			
origines		Déserts	Eaux et Forêts	Glaces	Océans
	Déserts	93.95	5.68	0.37	0
	Eaux et Forêts	13.84	83.80	2.34	0.02
	Glaces	0	8.63	91	0.37
	Océans	0	0.04	0.02	99.94

TAB. 3.5 – Table de confusion du classifieur en % après l’utilisation d’une AFD sur les paramètres des échos.

pour effectuer la classification. Pour cela, la moitié des signaux sont utilisés pour l’étape d’apprentissage, l’autre moitié est quant à elle utilisée pour le test de la méthode. Les résultats de classification illustrés dans le tableau 3.5 mettent en évidence une baisse des performances de classification. Ces résultats nous montrent que la classe « Eaux et Forêt » est la principale source des confusions. Contrairement au cas rencontré pour des données en 1Hz, la multiplication du nombre de classes ne permet pas l’amélioration des résultats. Malgré ce fait, nous pouvons tout de même parler d’une méthode aux performances acceptables.

3.3.5 Conclusion

Les résultats montrent que, si on travaille sur des paquets de 20 échos moyennés, parmi toutes les solutions envisagées au paragraphe 3.3.2, la meilleure solution consiste à faire la classification sur le vecteur issu de la projection complète par l’ACP du vecteur paramètre (range, elevation, peakiness, swl, sigma0). L’utilisation d’une AFD s’est avérée moins efficace à cause d’une variance intraclasse des « Glaces » largement supérieure à celles des autres classes, pénalisant ainsi l’optimisation des variances interclasses, lors du calcul des vecteurs propres, et donc la bonne séparation des classes. Toutefois, la séparation de la classe « Glaces » en deux sous-classes permet d’obtenir d’excellents résultats, comparables à ceux avec l’ACP en prétraitement.

Par contre, si on souhaite appliquer la classification écho par écho, les meilleurs résultats sont obtenus sur le vecteur de données issu de la projection sur les axes définis par l'AFD du vecteur paramètre (range, elevation, peakiness, kurtosis, skewness). Ce vecteur n'est pas exactement identique à celui utilisé dans la classification par paquet de 20 échos car nous ne travaillons plus avec des données moyennées. De ce fait, le domaine de chacune des classes devient plus important. Le vecteur de données initialement utilisé dans le cas de données moyennées s'est alors avéré moins pertinent et ce, quel que soit le prétraitement choisi.

3.4 Problème de classification PISTACH

3.4.1 Présentation du problème

Un autre problème de classification des formes d'onde altimétriques consiste à considérer des profils types ou « squelettes » d'écho, comme cela est décrit dans le projet PISTACH [Thi08]. Suite à une expertise, la figure 3.17 résume les formes retenues dans ce projet. Une telle classification permettrait d'adapter les algorithmes en fonction du type de « squelette » rencontré et ainsi obtenir le maximum d'information de l'ensemble des mesures réalisées par les satellites. Actuellement, la solution proposée est une classification par réseau de neurones [Thi08]. Une des contraintes fortes du réseau de neurones, est la nécessité de modifier la structure du réseau lorsque les conditions de mesure changent : définition du nombre de couche et de neurones par couche. Ce changement de structure peut s'avérer long et fastidieux puisque fait manuellement.

Dans cette partie nous proposons une méthode de classification en deux étapes. La première consiste à détecter les échos normaux de Brown (classe 1). Pour cela, nous testerons une approche par seuillage sur un certain nombre de paramètres, puis une approche basée sur la méthode du « One Class SVM ». Pour finir, les autres classes seront traitées en utilisant le classifieur bayésien sur un nombre variable de paramètres.

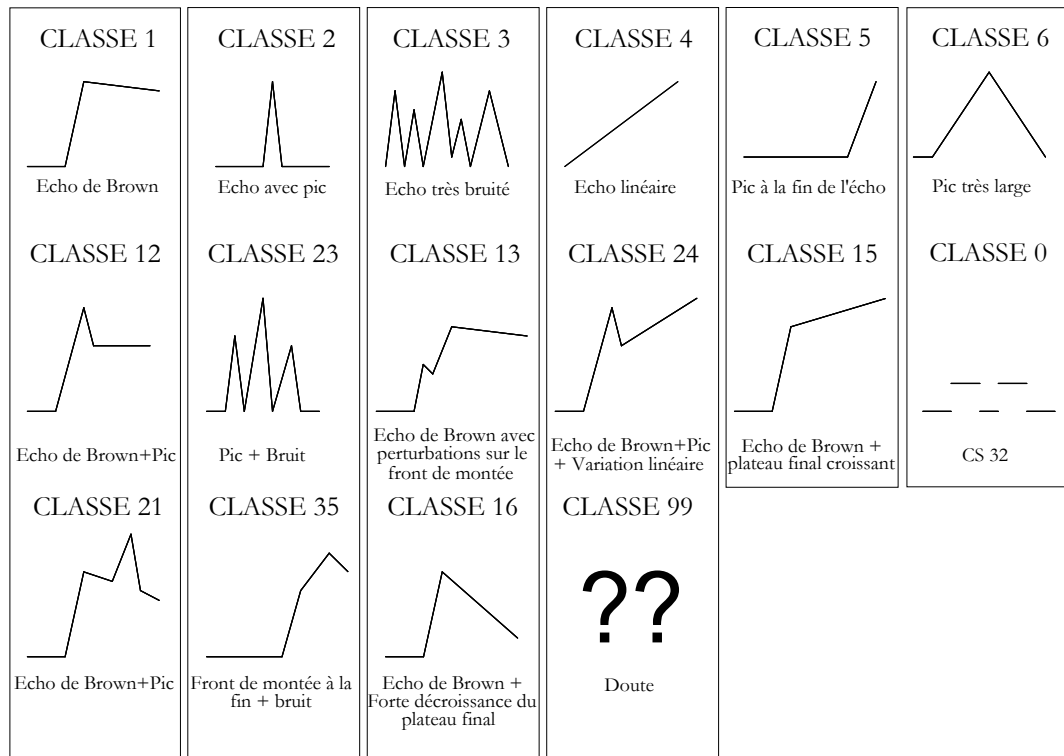


FIG. 3.17 – Schéma de principe pour la classification des formes des échos du projet PISTACH.

3.4.2 Détection d'anomalies

Les méthodes de détection d'anomalies permettent avant tout d'écarter une classe particulière, simple à identifier. Ces méthodes permettent d'accélérer la chaîne de traitement en écartant rapidement les échos de Brown qui correspondent ici à la première classe.

Méthode du seuil

Dans le cas précis de la détection d'anomalies, nous simplifions le problème présenté sur la figure 3.17. Les échos de Brown de la classe 1 constitueront la classe normale. Tous les autres signaux formeront la classe anomalie. Nous positionnant dans le cadre d'une

détection de signaux dits de Brown (la classe normale), nous procédons tout d'abord à l'estimation des paramètres de Brown sur l'ensemble des signaux d'apprentissage avec les algorithmes du maximum de vraisemblance à 3 puis 4 paramètres. Ainsi, pour chaque écho, nous disposons de l'erreur quadratique moyenne entre la forme d'onde et le modèle théorique le plus proche. Nous les appellerons MQE3 et MQE4.

Les estimations des densités de probabilité des paramètres MQE3 et MQE4 sont représentées sur les figures 3.18 et 3.19. On peut remarquer que les erreurs quadratiques des signaux de la classe normale sont relativement concentrées sur des faibles valeurs. Afin de fixer un seuil et d'obtenir une bonne classification, nous analysons les fonctions de répartition de ces erreurs quadratiques pour les paramètres MQE3 et MQE4 (voir figures 3.18 et 3.19). Nous choisissons alors un seuil $\gamma = 0.02$ nous permettant ainsi de bien classer environ 98% de nos signaux de la classe normale. Les tableaux 3.6 et 3.7

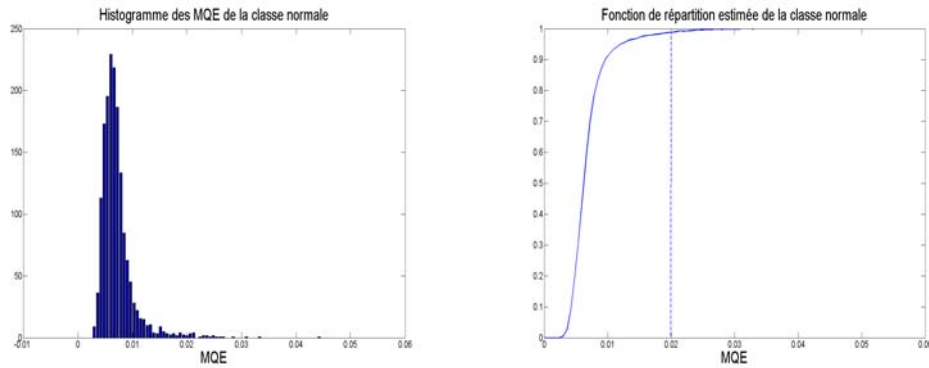


FIG. 3.18 – Histogramme et fonction de répartition estimée des MQE3s.

résument les résultats obtenus en utilisant ce seuil γ soit sur MQE3, soit sur MQE4. Malgré le choix d'un seuil approprié, les performances de classification des signaux de la classe anormale ne sont pas satisfaisantes.

Une analyse plus fine de ces résultats de classification nous a fait remarquer que malgré des résultats similaires, les erreurs d'affectation en fonction des MQE3 ou des MQE4 ne sont pas identiques. Nous avons donc décidé d'utiliser ces deux erreurs quadratiques

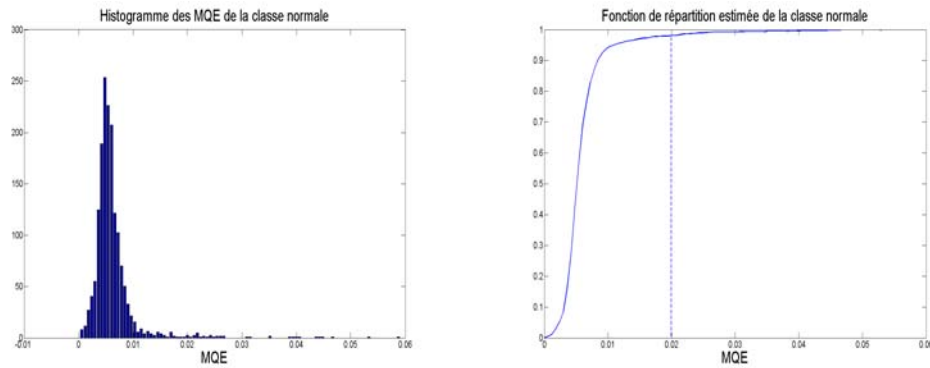


FIG. 3.19 – Histogramme et fonction de répartition estimée des MQE4s.

origines	affectations		
	MQE3	Brown	Non Brown
	Brown	98.70	1.30
	Non Brown	37.30	62.70

TAB. 3.6 – Table de confusion pour le seuillage des MQE3s.

origines	affectations		
	MQE4	Brown	Non Brown
	Brown	97.85	2.15
	Non Brown	28.10	71.90

TAB. 3.7 – Table de confusion pour le seuillage des MQE4s.

afin d'affiner la classification. La figure 3.20 illustre la répartition des signaux altimétriques dans l'espace erreurs (MQE3, MQE4). Combinant les informations apportées par ces erreurs quadratiques, nous avons déterminé visuellement un seuil $\gamma = 0.013$. Nous avons donc comparé les deux MQEs à ce seuil pour la classification. La table de confusion 3.8 expose les résultats obtenus. Bien sûr, le seuil γ ayant diminué, le pourcentage de signaux « Brown » correctement détectés a diminué : on passe de 98% à 93%, par rapport au seuillage sur l'une ou l'autre MQE. Mais l'utilisation conjointe des deux MQEs permet de diminuer globalement les erreurs de classification et d'augmenter le pourcentage de

signaux non Brown correctement classifiés. Ainsi l'utilisation conjointe des deux MQEs s'avère satisfaisante.

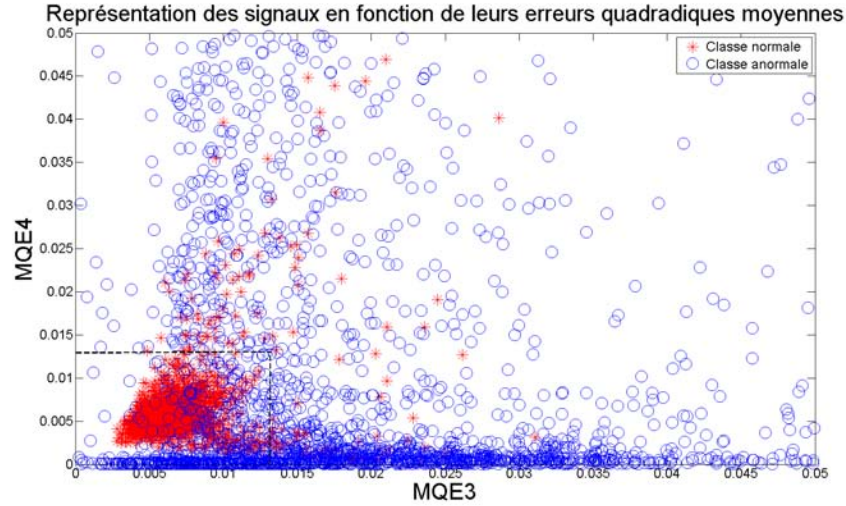


FIG. 3.20 – Représentations des signaux altimétriques en fonction de leurs MQEs.

origines	affectations	
	MQE	
	Brown	Non Brown
Brown	93.04	6.96
Non Brown	8.80	91.20

TAB. 3.8 – Table de confusion pour le seuillage des deux MQEs.

Cependant l'utilisation de ces deux MQEs peut s'avérer difficile à justifier car lors de la phase d'estimation des paramètres une seule méthode d'estimation est utilisée, et donc une seule MQE est disponible. Pour cette raison nous testerons aussi la détection d'anomalies sur des paramètres altimétriques issus du modèle de Brown. Selon l'algorithme d'estimation utilisé ($MLE3$ ou $MLE4$) notre vecteur paramètre sera composé de σ_0 (le coefficient de rétrodiffusion), l'Epoch une donnée altimétrique fonction de τ (écart de position en m d'une forme d'onde, telle qu'elle a été positionnée dans la fenêtre d'analyse lors de la phase d'estimation, et une porte de référence) et la hauteur moyenne des vagues

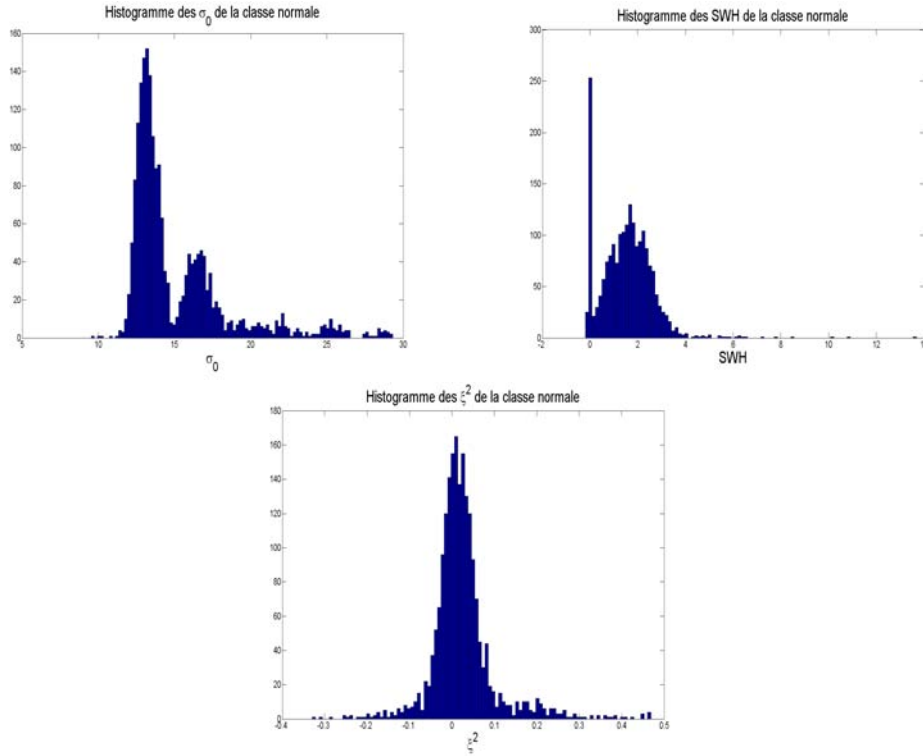


FIG. 3.21 – Histogrammes des paramètres de la classe normale.

swh (la hauteur moyenne des vagues) pour le *MLE3*, et du même vecteur paramètre complété par le dépointage de l'antenne ξ^2 (carré de la valeur du dépointage de l'antenne) pour le *MLE4*. De la même manière qu'avec les MQEs, l'analyse des histogrammes de la figure 3.21 nous permet de fixer des seuils afin de procéder à la classification. Les seuils choisis sont :

$$\begin{aligned}\sigma_0 &\geq 30, \\ \tau &\neq 0, \\ SWH &\geq 14, \\ \xi^2 &\geq 0.5 \text{ ou } \xi^2 \leq -0.35.\end{aligned}$$

Le paramètre Epoch nous informe sur la divergence de la phase d'estimation des paramètres :

- Epoch sera à zéro si la phase d'estimation a convergé,
- sinon elle prendra une très grande valeur pour signaler la divergence de la phase d'estimation.

Les matrices de confusions 3.9 et 3.10 nous permettent de constater que le choix de ces vecteurs nous mène à des méthodes aux performances similaires et très satisfaisantes.

origines	affectations		
	MQE	Brown	Non Brown
	Brown	100	0
	Non Brown	9.8	90.2

TAB. 3.9 – Table de confusion pour le seuillage des paramètres issus du *MLE3*.

origines	affectations		
	MQE	Brown	Non Brown
	Brown	100	0
	Non Brown	6.05	93.95

TAB. 3.10 – Table de confusion pour le seuillage des paramètres issus du *MLE4*.

Méthode du « One class SVM » sur les paramètres MQE3 et MQE4

L'utilisation d'un double seuil (sur MQE3 et MQE4) permet de créer une zone de décision, sur la figure 3.20, assimilable à un carré. En observant plus finement la figure, cette zone de décision peut être affinée. Dans ce problème à deux classes, nous pensons notamment à une forme non linéaire qu'il est possible d'obtenir en utilisant les algorithmes basés par exemple sur la méthode « One Class SVM » expliqués plus en détail dans la partie 3.2.4. Il s'agit d'une méthode d'apprentissage non supervisée (c.a.d utilisant exclusivement l'information des signaux de la classe normale), rendue supervisée avec l'emploi de la technique d'ajustement de noyau (le critère d'alignement) nécessitant les informations des signaux de la classe anormale.

Dans ce cas précis, nous disposons de deux classes : la classe normale (regroupant les signaux dits de Brown) et la classe anomalie (regroupant tous les autres signaux) construites à partir de la figure 3.17. Avant toute chose, nous effectuons un prétraitement sur notre base de données dans le but d'en exclure toutes les valeurs aberrantes. Plus précisément, les signaux dont la somme des MQEs est supérieure à 1 sont exclus. En effet, tous ces signaux rejetés de la base de données appartiennent à la classe anomalie. Leur classification est donc automatique.

Ensuite, dans le but de garantir la meilleure représentativité des deux classes, nos signaux d'apprentissage seront sélectionnés à partir des distributions bi-dimensionnelles des signaux de la classe normale et de la classe anomalie. De plus, nous considérons parmi ces signaux d'apprentissage qu'un certain nombre peut être exclu de la zone de décision finale en raison de la dispersion des signaux observables sur la figure 3.20. Pour cela, nous fixons le pourcentage d'« outliers » $\nu = 10\%$.

Nous commençons par utiliser 100 signaux de chaque classe afin d'obtenir les MQE3 et les MQE4 nécessaires à l'apprentissage des SVM. Le noyau choisi pour les SVM est un noyau gaussien pour lequel nous ferons varier la largeur c (ou variance). Nous testerons alors chaque noyau (caractérisé par sa largeur c) afin d'évaluer le critère d'alignement de chacun. Ainsi nous pourrions définir la variance optimale en vue d'une classification.

La figure 3.22 montre clairement l'évolution de ce critère d'alignement pour les différents noyaux évalués. Il existe un maximum global pour $c = 1.09 \cdot 10^{-5}$. Nous utilisons donc cette largeur de noyau dans le but de construire notre algorithme SVM. La figure 3.23 met en avant la zone de sélection obtenue dans ces conditions d'expérimentation. Les signaux de la classe normale sont effectivement bien séparés des autres signaux. Toutefois, le volume de la classe est faible, dû au faible nombre de signaux d'apprentissage et est de plus divisé en deux parties séparées. La table de confusion 3.11 met justement en avant ce problème lié à l'apprentissage. Les résultats sont obtenus en testant tous les signaux disponibles. Les anomalies sont détectées dans plus de 99% des cas, alors que la performance de classification des signaux de la classe normale reste modérée. Avec

un total de 1998 signaux appartenant à la classe normale et 6290 signaux à la classe anormale, nous arrivons à une erreur globale de classification de 5.85%.

origines	affectations		
	MQE	Brown	Non Brown
	Brown	78.73	21.27
	Non Brown	0.95	99.05

TAB. 3.11 – Table de confusion obtenue à partir du SVM avec 100 signaux d'apprentissage par classe.

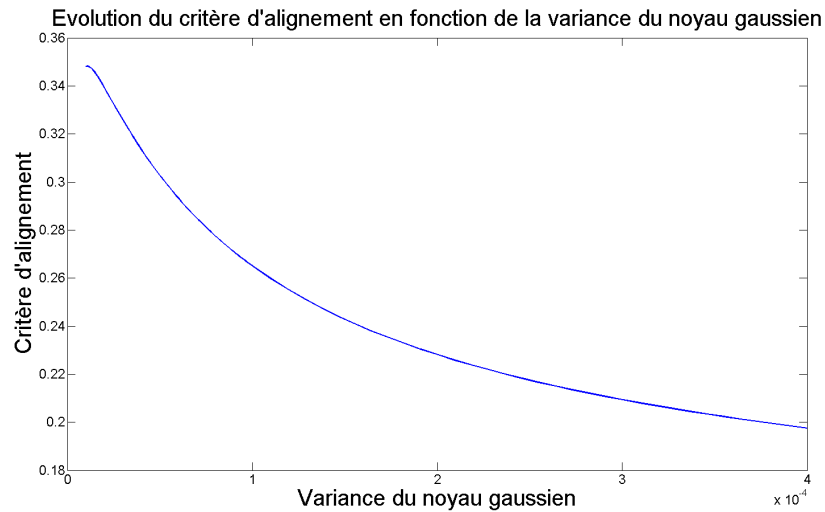


FIG. 3.22 – Évolution du critère d'alignement en fonction de la variance du noyau gaussien testé à partir de 100 signaux d'apprentissage par classe.

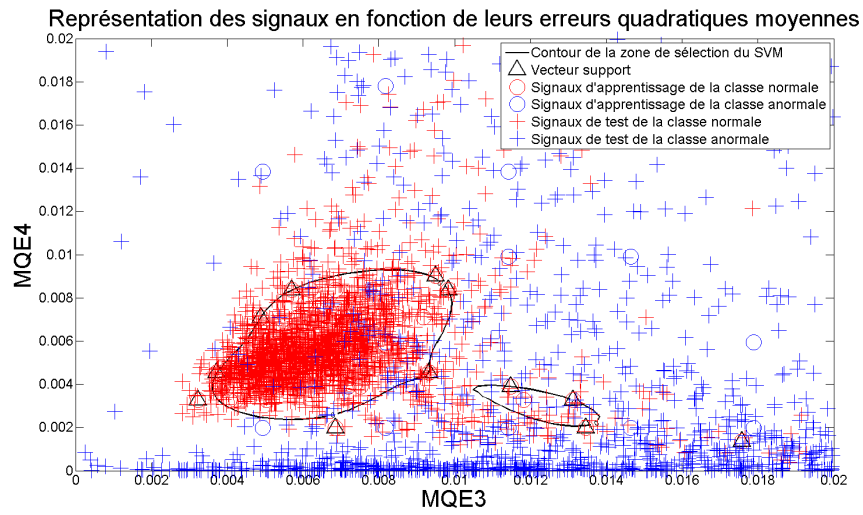


FIG. 3.23 – Résultat de la zone de séparation obtenue avec le SVM de critère d'alignement maximum (à partir de 100 signaux d'apprentissage).

Nous renouvelons donc l'expérimentation avec 500 signaux d'apprentissage pour chacune des classes. La figure 3.24 illustre la nouvelle progression du critère d'alignement en fonction de la largeur du noyau c . Le nouveau maximum global se situe à $c = 2.09.10^{-5}$. Utilisant cette largeur de noyau, la figure 3.25 illustre la nouvelle zone de sélection des classes. Contrairement au cas précédent, la zone obtenue est homogène, laissant supposer une plus grande marge de manœuvre pour les futurs signaux à tester. La table de confusion 3.12, montre qu'effectivement cette homogénéité permet d'améliorer les performances de classification. Avec un gain de presque 8% pour le signaux de la classe normale et une perte de seulement 1% pour la classe anormale, l'erreur globale est de 4.74%.

origines	affectations		
	MQE	Brown	Non Brown
	Brown	86.04	13.96
	Non Brown	1.81	98.19

TAB. 3.12 – Table de confusion obtenue à partir du SVM avec 500 signaux d'apprentissage par classe.

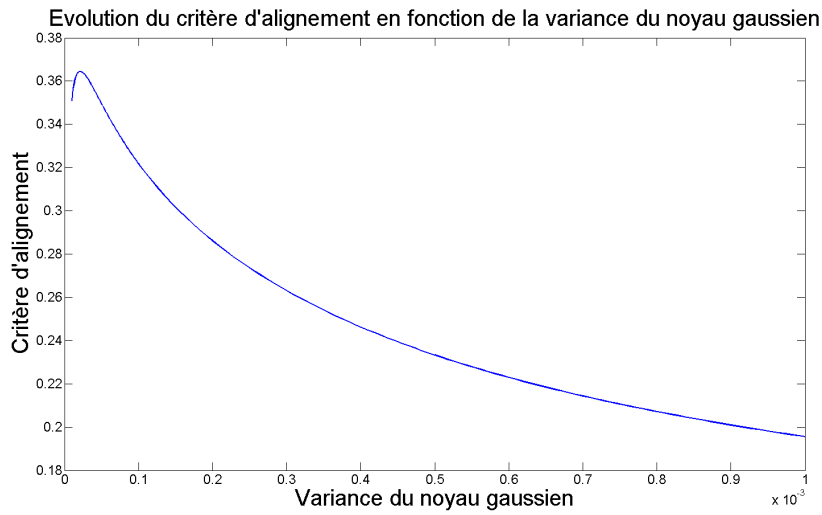


FIG. 3.24 – Évolution du critère d'alignement en fonction de la variance du noyau gaussien testé à partir de 500 signaux d'apprentissage par classe.

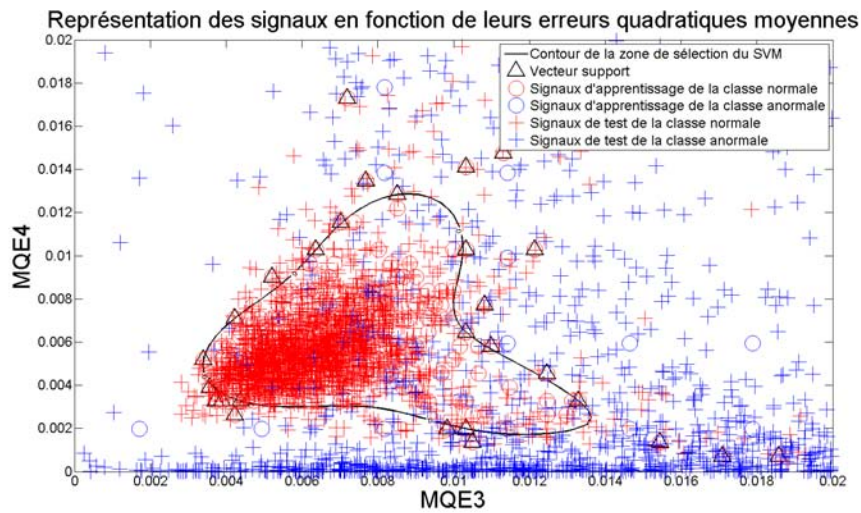


FIG. 3.25 – Résultat de la zone de séparation obtenue avec le SVM de critère d'alignement maximum (à partir de 500 signaux d'apprentissage).

Enfin, nous testons la méthode en utilisant 1000 signaux d'apprentissage par classe. La figure 3.26 affiche la progression du critère d'alignement et le maximum global de la fonction se trouve pour une largeur de noyau $c = 2.10^{-5}$. Ce maximum du critère d'alignement diffère très peu de la largeur du noyau obtenue avec 500 signaux d'apprentissage. Cette largeur ne changeant pas, seuls les vecteurs supports délimitant la zone de sélection vont être affectés. La figure 3.27 permet de mieux visualiser l'impact de l'augmentation du nombre de signaux d'apprentissage sur la zone de sélection finale. Les performances de classification ne sont que peu améliorées comme l'illustre la table de confusion 3.13. La classification relative à la classe normale est améliorée d'environ 4% tandis que celle relative à la classe anormale est altérée de presque 1%. Finalement, l'erreur globale descend à 4.48%.

origines	affectations		
	MQE	Brown	Non Brown
	Brown	90.19	9.81
	Non Brown	2.78	97.22

TAB. 3.13 – Table de confusion obtenue à partir du SVM avec 1000 signaux d'apprentissage par classe.

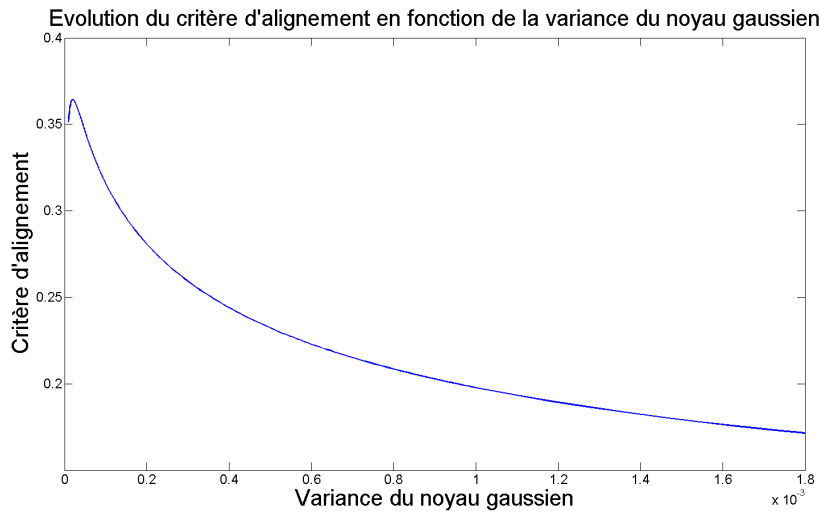


FIG. 3.26 – Évolution du critère d'alignement en fonction de la variance du noyau gaussien testé à partir de 1000 signaux d'apprentissage par classe.

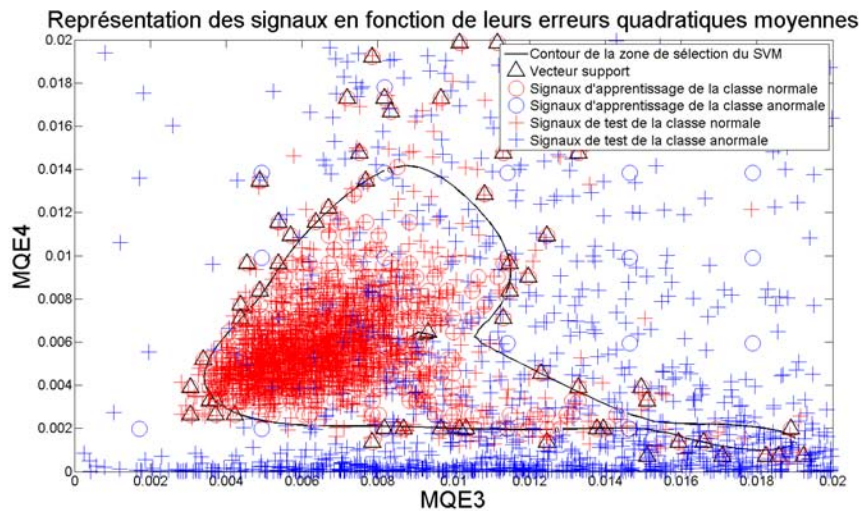


FIG. 3.27 – Résultat de la zone de séparation obtenue avec le SVM de critère d'alignement maximum (à partir de 1000 signaux d'apprentissage).

Méthode « One class SVM » sur des paramètres altimétriques

Dans le cas d'une détection d'anomalie avec un vecteur paramètre de dimension $q > 2$, il s'avère difficile de générer des données d'apprentissage à partir de leur distribution q -dimensionnelle. Pour cette raison, l'évaluation des performances de la détection d'anomalies sera effectuée différemment. La procédure utilisée se compose de quatre étapes :

- Tout d'abord, les données aberrantes sont mises de côté puisqu'elles correspondent à une divergence du tracker et ne peuvent être utilisées pour l'apprentissage.
- Ensuite, parmi les données restantes, 500 signaux de la classe normale et de la classe anomalie sont sélectionnés aléatoirement puis normalisés. Ces données nous permettent alors de définir la largeur de noyau c optimale pour l'apprentissage. Dans ce cas précis, la normalisation nous permet de tester des largeurs de noyau comprises entre 0.001 et 2.
- Les 500 signaux de la classe normale sont alors utilisés pour l'apprentissage de la fonction de décision de la méthode « One Class SVM » en utilisant la largeur de noyau c définie précédemment et une proportion d'« outliers » $\nu = 1\%$.
- Enfin, nous testons la classification sur les 18305 signaux restant de notre base de données (données aberrantes incluses) normalisés en fonction de la moyenne et de l'écart type définis dans la phase précédente.

Cette procédure est alors itérée 100 fois afin de pouvoir étudier les statistiques des performances. Les matrices de confusions 3.14 et 3.15 nous permettent de constater de très bonnes performances avec l'utilisation des paramètres altimétriques. De plus ces résultats sont statistiquement stables puisque l'on constate des écarts types et donc des variances très faibles.

origines	affectations			
	MQE	Brown	Non Brown	σ
	Brown	96.72	3.28	0.83
	Non Brown	3.34	96.66	0.22

TAB. 3.14 – Table de confusion moyenne obtenue avec les paramètres du *MLE3*.

origines	affectations			σ
	MQE	Brown	Non Brown	
	Brown	95.55	4.45	
	Non Brown	1.09	98.91	0.17

TAB. 3.15 – Table de confusion moyenne obtenue avec les paramètres du *MLE4*.

Conclusion

Dans cette partie nous avons présenté les performances de deux méthodes de détection d'anomalie utilisables comme prétraitement au problème illustré sur la figure 3.17 :

- la détection d'anomalie par seuil,
- les machines à vecteurs supports à une classe (« One Class Support Vector Machine » en anglais, OC-SVM).

L'utilisation d'un seuil pour la première méthode permet de mettre en place une détection d'anomalies rapidement et simplement. Seulement pour cela, l'utilisation de l'ensemble des signaux de la base de données est nécessaire afin de fixer un seuil selon nos observations. Les performances de classification d'une telle approche sont très satisfaisantes puisque plus de 90% des signaux des deux classes sont bien détectés.

Pour des performances équivalentes, il est toutefois possible de mettre en place une méthode automatique d'apprentissage en utilisant la méthode *One-Class SVM* combinée avec une méthode de sélection de noyau : le critère d'alignement. Cette approche permet d'utiliser moins de signaux d'apprentissage que la méthode par seuil. L'apprentissage se fait de manière automatique au détriment d'un temps de calcul largement supérieur. Ceci dit, les résultats de classification avec la méthode des SVMs sont comparables à ceux obtenus avec un seuil, la détection d'anomalies, au sens propre du terme, étant tout de même légèrement meilleure.

3.4.3 Classification bayésienne à partir des paramètres

La détection d'anomalies présentée au paragraphe précédent nous a permis d'isoler la classe 1 des signaux dits de Brown des autres signaux. Dans une deuxième étape, nous nous concentrons sur la classification des autres signaux, considérés comme non Brown. Nous analyserons dans cette partie les résultats obtenus en utilisant la méthode du classifieur bayésien à partir des distances de Mahalanobis et proposerons une alternative aux différentes classes proposées dans le projet PISTACH, illustrées sur la figure 3.17.

Prétraitement des échos

Dans cette partie, nous considérons que les signaux non Brown ne nécessitent pas l'utilisation des erreurs quadratiques moyennes d'estimation du *MLE* car ces signaux ne suivent pas le comportement décrit par le modèle de Brown. Nous utiliserons donc des données issues de la phase d'estimation des paramètres auxquelles nous ajouterons des paramètres liés à la forme de l'écho.

Les données issues de la phase d'estimation des paramètres concernent la latitude (LAT) et la longitude (LON), le coefficient de rétrodiffusion (SIG0), un coefficient lié à la hauteur de vague moyenne (SIGC) et enfin la hauteur de vague moyenne (SWH).

De plus, nous déterminons pour chaque écho, sa valeur maximum (V_{MAX}) et sa valeur moyenne (V_{MOY}) nous permettant de définir le peakiness de l'écho [Thi08] :

$$peakiness = \frac{V_{MAX}}{V_{MOY}}. \quad (3.35)$$

Ensuite nous calculons des coefficients relatifs aux moments d'ordre 2 à 4, qui sont respectivement la variance, le skewness et le kurtosis :

$$skewness = \frac{\mathbb{E}[(x - \mathbb{E}[x])^3]}{\left[\sqrt{\mathbb{E}[(x - \mathbb{E}[x])^2]}\right]^3}, \quad (3.36)$$

$$kurtosis = \frac{\mathbb{E}[(x - \mathbb{E}[x])^4]}{[\mathbb{E}[(x - \mathbb{E}[x])^2]]^2}. \quad (3.37)$$

Enfin, pour des échos composés de 104 portes, nous calculons la pente entre les portes 1 et 60, appelée MONTEE, et entre les portes 40 à 104 que l'on nommera ATTITUDE.

Problématique initiale : PISTACH

Avant toute chose, nous procédons à une AFD afin de déterminer le pouvoir discriminant des divers paramètres utilisés. La figure 3.28 nous permet de se rendre compte que les 3 paramètres les plus discriminants s'avèrent être des paramètres liés à l'écho plutôt qu'aux données issues de la phase d'estimation des paramètres. La variance, le kurtosis ainsi que le skewness sont les principaux paramètres optimisant la séparation des classes.

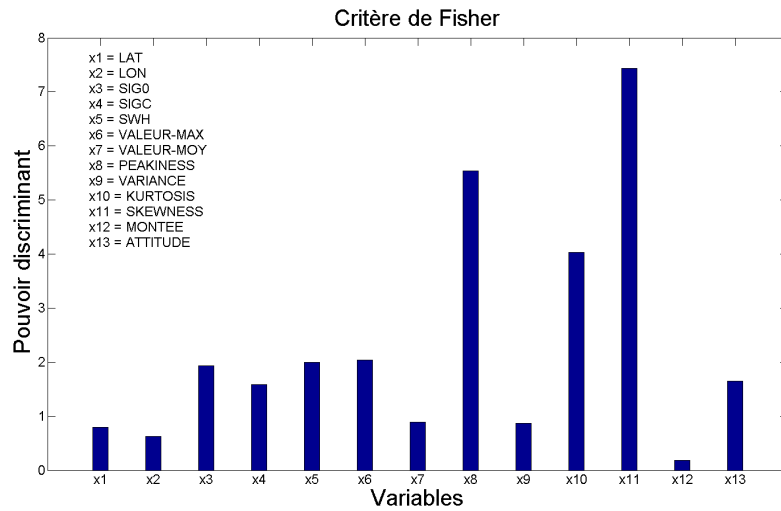


FIG. 3.28 – Critère de Fisher des variables.

La méthode de classification choisie pour ce cas précis, est le classifieur bayésien utilisant la distance de Mahalanobis. Disposant de peu de signaux par classe, la méthode d'évaluation de notre classification se fera par la technique du « Leave One Out ». Plus précisément, avec un total de N signaux, nous effectuerons une AFD et l'apprentissage

sur $N - 1$ signaux et classerons le signal restant, et ce pour la totalité de notre base de données.

La table de confusion 3.16 expose les résultats de classification obtenue. On remarque des performances acceptables pour l'ensemble des classes avec des pourcentages de bonne classification variant entre 80% et 100%. Seules les classes 6 et 16 connaissent des résultats assez moyens avec respectivement un taux de réussite de 65% et 76%.

affectations														
origines	Classes	2	3	4	5	6	12	13	15	16	21	23	24	35
	2	98	0	0	2	0	0	0	0	0	0	0	0	0
	3	0	86	2	2	0	8	0	2	0	0	0	0	0
	4	0	3	88	3	0	3	3	0	0	0	0	0	0
	5	0	0	0	100	0	0	0	0	0	0	0	0	0
	6	0	0	0	2	65	33	0	0	0	0	0	0	0
	12	0	0	0	12	0	83	0	0	0	3	0	2	0
	13	0	0	0	0	0	7	82	2	0	7	0	2	0
	15	0	0	0	0	0	2	11	85	0	0	0	2	0
	16	0	0	0	0	0	15	2	0	76	0	2	5	0
	21	0	0	0	0	0	10	4	0	0	86	0	0	0
	23	0	0	0	2	0	2	0	0	0	0	96	0	0
	24	0	0	0	3	0	8	0	0	2	0	0	87	0
	35	0	0	0	7	0	0	0	0	0	0	0	0	93

TAB. 3.16 – Table de confusion en utilisant le classifieur bayésien avec les distances de Mahalanobis sur 13 classes.

Regroupement de classes

Compte-tenu des résultats obtenus, l'observation des échos mal classés nous conduit à reconsidérer le découpage en classes proposé sur la figure 3.17. L'AFD nous permet de définir six grands groupes d'échos :

- Les « échos presque Brown » qui correspondent à des formes d’onde qui par leurs aspects rappellent fortement un écho de Brown au bruit près. Il s’agit des classes 13, 15 et 21 regroupées dans une classe notée R_1 .
- Les « pics » de la classe 2 sont conservés dans une classe notée R_2 .
- Les échos vraiment très proches d’un simple bruit (les classes 3 et 23) sont réunis dans la classe R_3 .
- Les signaux linéaires représentés par les classes 4 et 24 sont regroupés dans la classe R_4 .
- Les formes d’onde dont l’énergie se situe en fin d’écho : les classes 5 et 35 constituent la classe R_5 .
- Enfin les « larges pics » des classes 6, 12 et 16 sont placés dans la classe R_6 .

La figure 3.29 illustre le nouveau découpage proposé.

Nous appliquons alors le classifieur bayésien utilisant la distance de Mahalanobis sur ces six classes. La table de confusion 3.17 nous permet d’observer des performances acceptables pour 4 des 6 classes. La classe R_4 connaît la plus mauvaise performance avec un peu de moins de 50% des signaux bien classés. Ce regroupement de classe devait proposer un problème de classification plus simple que celui proposé avec les premières classes, mais les résultats de classification ne sont pas concluants et ce même en utilisant une ACP en prétraitement.

		affectations					
origines	Regroupement	R_1	R_2	R_3	R_4	R_5	R_6
	R_1	83	0	0	7	4	6
	R_2	0	98	0	0	2	0
	R_3	0	0	85	3	9	3
	R_4	4	0	16	48	30	2
	R_5	0	0	2	0	98	0
	R_6	8	0	5	0	19	68

TAB. 3.17 – Table de confusion en utilisant le classifieur bayésien avec les distances de Mahalanobis sur 6 regroupements.

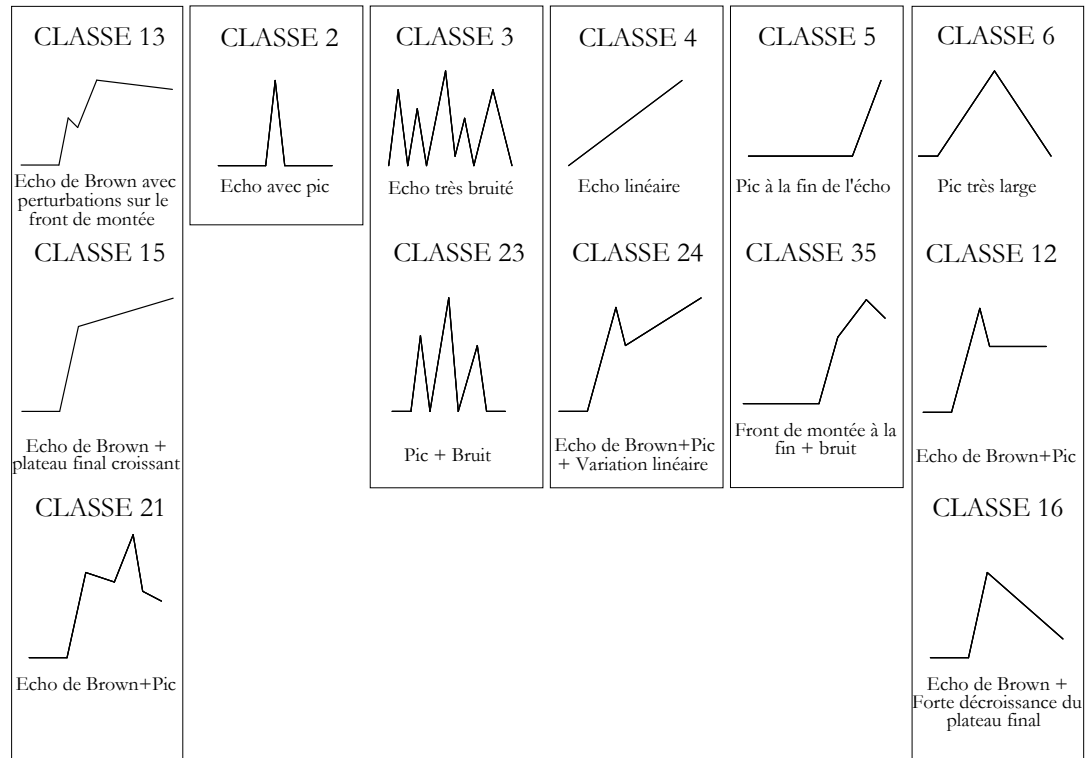


FIG. 3.29 – Nouveau schéma de principe pour la classification des formes des échos.

3.4.4 Conclusion

Dans cette partie nous avons proposé une solution pour le problème de classification abordé dans le projet PISTACH. Notre méthode de classification est en fait une chaîne de traitement composée de plusieurs étapes.

Dans un premier temps nous appliquons une détection d'anomalies pour séparer les signaux dits de Brown des autres signaux qualifiés d'anomalies. Les résultats des expérimentations montrent que la méthode du « One Class SVM » combinée avec le critère d'alignement permet d'obtenir de très bons résultats. En comparaison avec la méthode de seuillage, l'utilisation des SVMs permet une souplesse d'utilisation supplémentaire grâce aux différents choix possibles que sont :

- le choix du pourcentage d'« outliers » ν permettant de définir un taux d'erreur de classification,
- Le nombre de signaux utilisés pour l'apprentissage fortement lié aux nombres de vecteurs support définissant la zone de décision.

Les signaux dits de Brown ayant été écartés de la classification, les 13 classes restantes sont traitées en utilisant le classifieur bayésien. Les résultats montrent qu'il s'agit globalement d'une méthode efficace permettant de classer correctement presque la totalité des signaux.

Dans le chapitre suivant de cette thèse, nous proposons une nouvelle modélisation des formes d'onde intégrant la présence de pics. Nous espérons ainsi extraire des informations utiles des signaux altimétriques non « Browniens ».

CHAPITRE 4

Formes d'onde altimétriques non-océaniques

Sommaire

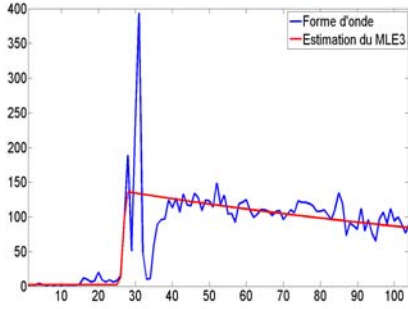
4.1	Introduction	109
4.2	Modélisation proposée	111
4.3	Méthode du maximum de vraisemblance	113
4.4	Les performances	115
4.5	Extension à la présence de plusieurs pics	126
4.6	Conclusion	128

4.1 Introduction

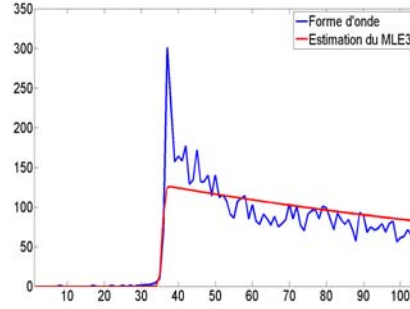
Dans la partie précédente, nous avons présenté un projet consacré à la classification des formes d'onde selon leurs profils types répartis dans 14 catégories illustrées sur la figure 3.17. Actuellement toutes les classes, hormis la classe 1 contenant des échos dits de Brown, conduisent à un rejet des signaux, et donc ne sont pas exploités. Pourtant, certaines de ces classes regroupent des échos proches des signaux altimétriques océaniques avec un ou plusieurs pics comme perturbations supplémentaires. C'est le cas des classes 12, 13 et 21 illustrées sur la figure 3.17. Ces signaux viennent d'échos retour de surfaces non océaniques (glaces, terres, zones côtières) pour lesquelles le modèle de Brown n'est pas adapté. En effet, dans ce contexte particulier, il s'avère que l'estimateur du maximum de vraisemblance approché actuellement utilisé, basé sur le modèle de Brown, ne permet pas de conserver des estimations fiables et précises.

Dans cette partie nous proposons donc d'utiliser une nouvelle modélisation des formes d'onde altimétriques corrompues par des pics. Pour commencer, nous divisons en trois catégories les différents types de manifestations des pics en fonction de leurs localisations sur l'écho : avant le front de montée, sur le front de montée et sur le plateau final. La figure 4.1 illustre ces trois catégories en superposant les échos perturbés extraits de données réelles du projet PISTACH [Thi08] ainsi que l'estimation obtenue en utilisant l'algorithme du *MLE3* (prenant compte les trois principaux paramètres altimétriques : P_u , τ et swh).

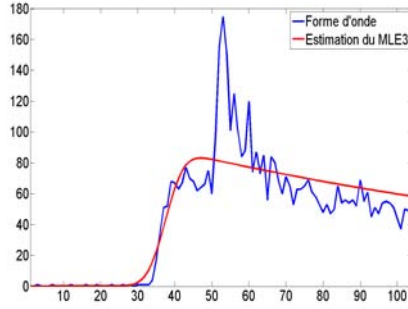
La figure 4.1(a) nous permet d'observer l'imprécision des estimations de τ et swh lorsque le pic se situe avant le front de montée. Ceci s'explique par le fait que l'estimateur *MLE3* capture le pic pensant qu'il s'agit du front de montée, induisant ainsi une estimation de ces paramètres à partir du pic et non du front de montée lui-même. On comprend donc bien que la qualité des estimations des paramètres altimétriques est directement liée aux caractéristiques du pic (amplitude et largeur). Dans la deuxième catégorie, illustrée sur la figure 4.1(b), le front de montée de l'écho et le pic se confondent. L'estimateur *MLE3* capture correctement le front de montée et donc effectue des estimations de τ et de swh relativement précises. Cependant, la puissance du signal P_u est directement affectée par l'amplitude du pic et donc constamment surévaluée. Enfin, dans la dernière catégorie, figure 4.1(c), l'ensemble des caractéristiques du pic (position, amplitude et largeur) pénalise l'algorithme *MLE3* dont les estimations ne sont plus fiables. L'ensemble de ces problèmes d'estimation pour des échos perturbés par des pics nous a donc mené à utiliser une modélisation spécifique à ces catégories de signaux altimétriques souvent rencontrés lors de mesures proches des côtes [GEVQ⁺ar] et qui pour le moment restent inexploités.



(a) Pic avant le front de montée.



(b) Pic sur le front de montée.



(c) Pic sur le plateau final.

FIG. 4.1 – Les trois types de manifestation d'un pic dans une forme d'onde pour des signaux extraits du projet PISTACH [Thi08].

4.2 Modélisation proposée

Une forme d'onde altimétrique classique (mesure effectuée en milieu hauturier) se modélise en utilisant la forme développée de Hayne s_k [Hay80] issue de la double convolution proposée par Brown [Bro77]. Seulement cette modélisation des échos n'est pas appropriée pour les signaux étudiés dans cette partie. Nous proposons ici d'utiliser un modèle mathématique simple qui nous autorise à considérer les échos altimétriques comme une superposition d'un écho de Hayne et d'une somme de pics gaussiens :

$$\tilde{s}_k = s_k + p_k, \quad k = 1, \dots, K, \quad (4.1)$$

avec K le nombre de points du signal altimétrique, s_k issu du modèle de Hayne et p_k le terme modélisant la présence de pics. Supposant la présence d'un nombre q de pics, le modèle que nous proposons d'étudier (proche de celui de [GEVQ⁺ar]) s'écrit :

$$p_k = \sum_{i=1}^q A_i \exp \left[-\frac{1}{2} \left(\frac{k - k_i}{v_i} \right)^2 \right]. \quad (4.2)$$

Chaque pic est modélisé par une gaussienne d'amplitude A_i , de position k_i et de largeur v_i . Pour une meilleure lisibilité, nous ferons à présent la distinction entre les paramètres altimétriques du modèle s_k et les paramètres liés à la modélisation des pics p_k :

- $\boldsymbol{\theta}_b = (P_u, \tau, \text{swh})^T$ représente le vecteur des paramètres altimétriques du modèle de Hayne $\mathbf{s} = [s_1, \dots, s_K]$,
- $\boldsymbol{\theta}_p = (\mathbf{A}^T, \mathbf{k}^T, \mathbf{v}^T)^T$ est le vecteur des paramètres des pics modélisés par $\mathbf{p} = [p_1, \dots, p_K]$, avec $\mathbf{A} = (A_1, \dots, A_q)^T$, $\mathbf{k} = (k_1, \dots, k_q)^T$ et $\mathbf{v} = (v_1, \dots, v_q)^T$.

Comme nous l'avons déjà fait dans le cas de signaux altimétriques océaniques, nous considérons deux sources de bruits affectant les mesures :

- le bruit thermique noté N_t (lié aux instruments à bord du satellite) qui est un bruit additif,
- le bruit multiplicatif dû au phénomène du speckle noté R_k modélisé par une loi gamma, $R_k \sim \mathcal{G}(L, L)$, L étant le nombre de signaux sommés de manière incohérente afin de réduire l'effet du bruit de « speckle ».

La forme d'onde altimétrique $\mathbf{y} = (y_1, \dots, y_K)$ est donc définie par :

$$y_k = (\tilde{s}_k + N_t) R_k. \quad (4.3)$$

Si $\boldsymbol{\theta} = (\boldsymbol{\theta}_b^T, \boldsymbol{\theta}_p^T)^T$ est le vecteur paramètre du modèle (4.1), la log-vraisemblance des données $\mathbf{y} = (y_1, \dots, y_K)$ du modèle (4.3) s'écrit :

$$\ln [f(\mathbf{y}|\boldsymbol{\theta})] = \text{cste} + (L-1) \sum_{k=1}^K \ln(y_k) - L \sum_{k=1}^K \ln(\tilde{s}_k) - L \sum_{k=1}^K \frac{y_k}{\tilde{s}_k}. \quad (4.4)$$

4.3 Méthode du maximum de vraisemblance

À partir de la méthode initialement développée par [Dum85] et détaillée en Annexe B, nous proposons une méthode d'estimation du maximum de vraisemblance appliquée au modèle (4.3) de vecteur paramètre $\boldsymbol{\theta} = (\boldsymbol{\theta}_b^T, \boldsymbol{\theta}_p^T)^T$. Cette méthode cherche à minimiser une fonction coût en utilisant la méthode de Quasi-Newton. Reprenant la notation de l'Annexe B, nous désirons maximiser la log-vraisemblance (4.4), de manière équivalente minimiser la fonction coût \tilde{C} définie par :

$$\tilde{C} = -\ln[f(\mathbf{y}|\boldsymbol{\theta})]. \quad (4.5)$$

Nous considérons le vecteur paramètre $\boldsymbol{\theta}_b$ comme un vecteur à m composantes ($m = 3, 4, 5$ suivant le modèle utilisé pour s_k) et le vecteur $\boldsymbol{\theta}_p = (\mathbf{A}^T, \mathbf{k}^T, \mathbf{v}^T)^T$ un vecteur composé de $3q$ composantes. Par simplicité, nous notons $\{\boldsymbol{\theta}_{b,i}\}_{i=1,\dots,m}$ et $\{\boldsymbol{\theta}_{p,j}\}_{j=1,\dots,3q}$ les composantes des vecteurs $\boldsymbol{\theta}_b$ et $\boldsymbol{\theta}_p$.

Ainsi le gradient de la fonction coût \tilde{C} s'écrit :

$$\nabla \tilde{C} = \begin{pmatrix} \frac{-\partial \ln[f(\mathbf{y}|\boldsymbol{\theta})]}{\partial \theta_{b,1}} \\ \vdots \\ \frac{-\partial \ln[f(\mathbf{y}|\boldsymbol{\theta})]}{\partial \theta_{b,m}} \\ \frac{-\partial \ln[f(\mathbf{y}|\boldsymbol{\theta})]}{\partial \theta_{p,1}} \\ \vdots \\ \frac{-\partial \ln[f(\mathbf{y}|\boldsymbol{\theta})]}{\partial \theta_{p,3q}} \end{pmatrix} = L \begin{pmatrix} \sum_{k=1}^K \frac{\tilde{s}_k - y_k}{\tilde{s}_k^2} \frac{\partial \tilde{s}_k}{\partial \theta_{b,1}} \\ \vdots \\ \sum_{k=1}^K \frac{\tilde{s}_k - y_k}{\tilde{s}_k^2} \frac{\partial \tilde{s}_k}{\partial \theta_{b,m}} \\ \sum_{k=1}^K \frac{\tilde{s}_k - y_k}{\tilde{s}_k^2} \frac{\partial \tilde{s}_k}{\partial \theta_{p,1}} \\ \vdots \\ \sum_{k=1}^K \frac{\tilde{s}_k - y_k}{\tilde{s}_k^2} \frac{\partial \tilde{s}_k}{\partial \theta_{p,3q}} \end{pmatrix}. \quad (4.6)$$

De par la modélisation théorique d'un écho en présence de pic (4.1), nous obtenons $\forall i \in (1, \dots, m)$ et $\forall j \in (1, \dots, 3q)$:

$$\begin{aligned} \frac{\partial \tilde{s}_k}{\partial \theta_{b,i}} &= \frac{\partial s_k}{\partial \theta_{b,i}}, \\ \frac{\partial \tilde{s}_k}{\partial \theta_{p,j}} &= \frac{\partial p_k}{\partial \theta_{p,j}}. \end{aligned}$$

Nous définissons alors les matrices $\tilde{\mathbf{D}} = (\tilde{D}_k)_{k=1,\dots,K}$, $\tilde{\mathbf{B}} = (\tilde{B}_{ik})_{\substack{i=1,\dots,m \\ k=1,\dots,K}}$ et $\tilde{\mathbf{P}} = (\tilde{P}_{jk})_{\substack{j=1,\dots,3q \\ k=1,\dots,K}}$ telles que :

$$\begin{aligned}\tilde{D}_k &= \frac{\tilde{s}_k - y_k}{\tilde{s}_k}, \\ \tilde{B}_{ik} &= \frac{\partial s_k}{\tilde{s}_k \partial \theta_{b,i}}, \\ \tilde{P}_{jk} &= \frac{\partial p_k}{\tilde{s}_k \partial \theta_{p,j}}.\end{aligned}$$

Le gradient de la fonction coût devient donc :

$$\nabla \tilde{C} = L \begin{pmatrix} \tilde{\mathbf{B}} \\ \tilde{\mathbf{P}} \end{pmatrix} \tilde{\mathbf{D}}^T. \quad (4.7)$$

De la même manière, la hessienne $\tilde{\mathbf{H}}$ est de la forme :

$$\tilde{\mathbf{H}} = \begin{bmatrix} \begin{bmatrix} \frac{\partial^2 \tilde{C}}{\partial \theta_{b,1} \partial \theta_{b,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{b,1} \partial \theta_{b,m}} \\ \vdots & & \vdots \\ \frac{\partial^2 \tilde{C}}{\partial \theta_{b,m} \partial \theta_{b,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{b,m} \partial \theta_{b,m}} \end{bmatrix} & \begin{bmatrix} \frac{\partial^2 \tilde{C}}{\partial \theta_{b,1} \partial \theta_{p,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{b,1} \partial \theta_{p,3q}} \\ \vdots & & \vdots \\ \frac{\partial^2 \tilde{C}}{\partial \theta_{b,m} \partial \theta_{p,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{b,m} \partial \theta_{p,3q}} \end{bmatrix} \\ \begin{bmatrix} \frac{\partial^2 \tilde{C}}{\partial \theta_{p,1} \partial \theta_{b,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{p,1} \partial \theta_{b,m}} \\ \vdots & & \vdots \\ \frac{\partial^2 \tilde{C}}{\partial \theta_{p,3q} \partial \theta_{b,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{p,3q} \partial \theta_{b,m}} \end{bmatrix} & \begin{bmatrix} \frac{\partial^2 \tilde{C}}{\partial \theta_{p,1} \partial \theta_{p,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{p,1} \partial \theta_{p,3q}} \\ \vdots & & \vdots \\ \frac{\partial^2 \tilde{C}}{\partial \theta_{p,3q} \partial \theta_{p,1}} & \cdots & \frac{\partial^2 \tilde{C}}{\partial \theta_{p,3q} \partial \theta_{p,3q}} \end{bmatrix} \end{bmatrix}, \quad (4.8)$$

qui s'exprime alors sous forme de blocs :

$$\tilde{\mathbf{H}} = L \begin{pmatrix} \tilde{\mathbf{B}}\tilde{\mathbf{B}}^T & \tilde{\mathbf{B}}\tilde{\mathbf{P}}^T \\ \tilde{\mathbf{P}}\tilde{\mathbf{B}}^T & \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \end{pmatrix}. \quad (4.9)$$

Finalement, la méthode itérative de Quasi-Newton, rappelée dans (B.3) et en posant $\Psi = 1$, d'estimation du vecteur paramètre $\boldsymbol{\theta}$ devient :

$$\boldsymbol{\theta}_{(t+1)} = \boldsymbol{\theta}_{(t)} - \begin{pmatrix} \tilde{\mathbf{B}}\tilde{\mathbf{B}}^T & \tilde{\mathbf{B}}\tilde{\mathbf{P}}^T \\ \tilde{\mathbf{P}}\tilde{\mathbf{B}}^T & \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\mathbf{B}} \\ \tilde{\mathbf{P}} \end{pmatrix} \tilde{\mathbf{D}}^T. \quad (4.10)$$

Le passage de la méthode originale rappelée en Annexe B à un estimateur *MLE3* en présence de pics, que nous appellerons *MLE3 Peak*, se fait finalement en insérant les dérivées partielles des pics dans le gradient et dans la hessienne de la fonction coût selon les équations (4.7) et (4.9). Nous avons alors $\forall i \in (1, \dots, q)$:

$$\frac{\partial \tilde{s}_k}{\partial A_i} = \exp \left[-\frac{1}{2} \left(\frac{k - k_i}{v_i} \right)^2 \right], \quad (4.11)$$

$$\frac{\partial \tilde{s}_k}{\partial k_i} = \frac{A_i (k - k_i)}{v_i^2} \exp \left[-\frac{1}{2} \left(\frac{k - k_i}{v_i} \right)^2 \right], \quad (4.12)$$

$$\frac{\partial \tilde{s}_k}{\partial v_i} = \frac{A_i (k - k_i)^2}{v_i^3} \exp \left[-\frac{1}{2} \left(\frac{k - k_i}{v_i} \right)^2 \right]. \quad (4.13)$$

4.4 Les performances

Afin d'étudier les performances de l'estimateur proposé *MLE3 Peak*, nous choisissons dans un premier temps de s'intéresser uniquement à la version « mono-pic » de cet estimateur dans lequel $q = 1$.

Les performances de cet estimateur seront comparées à :

- d'une part l'estimateur actuellement utilisé, *MLE3*, qui est une *approximation* du maximum de vraisemblance (voir B.3) et *qui ne prend pas en compte la présence de pic* dans la forme d'onde,
- d'autre part à un estimateur que nous qualifierons de référence. Cet estimateur permet d'obtenir l'estimateur « exact » du maximum de vraisemblance du modèle avec pics (4.4). Il utilise la méthode d'optimisation de *Nelder-Mead* [NM65]. Il s'agit d'un algorithme non-linéaire utilisant le concept de simplex et dont la convergence vers un minimum est étudiée dans [LRWW98]. Cette méthode est choisie à titre de référence car son utilisation dans l'estimation des paramètres altimétriques de manière opérationnelle serait trop coûteuse en calculs. Rappelons que l'estimateur que nous proposons, détaillé dans le paragraphe 4.3 correspond à une *approximation*

du maximum de vraisemblance, au même titre que l'estimateur *MLE3*, mais dans lequel nous prenons en compte la présence de pics.

L'évaluation des performances des trois algorithmes (*MLE3*, *MLE3 Peak* et *Nelder-Mead*) se fait en comparant les erreurs quadratiques moyennes (*Mean Square Error*, *MSE*). Pour cela, nous réalisons 100 estimations de formes d'onde synthétiques de mêmes paramètres P_u et τ perturbées par des réalisations de bruit différentes, et ce pour différentes valeurs de swh. Plus précisément, les deux paramètres altimétriques P_u et τ ont été fixés comme suit :

$$\begin{aligned} P_u &= 130, \\ \tau &= 31, \end{aligned}$$

et le paramètre swh varie dans l'intervalle $[0, 13\text{m}]$. Nous étudions les performances des algorithmes selon la position du pic d'amplitude $A_1 = 400$ et de variance $v_1 = 1$.

4.4.1 Pic avant le front de montée

La figure 4.2 est un exemple de signal altimétrique $\theta_b = (P_u, \tau, \text{swh}) = (130, 31, 2)$ altéré par un pic se situant avant le front de montée $\theta_p = (A_1, k_1, v_1) = (400, 26, 1)$. Nous pouvons constater sur la figure 4.2(a) que le *MLE3* confond le pic et le front de montée de la partie « altimétrique » de l'écho. Il en résulte des estimations de τ et swh erronées puisque calculées à partir du pic. Cependant, l'algorithme *MLE3 Peak*, figure 4.2(c), permet d'effectuer la distinction entre le pic et le front de montée, ce qui est en accord avec le résultat obtenu avec la fonction de référence *Nelder-Mead*, illustré figure 4.2(b).

Lorsque nous comparons en terme d'erreur quadratique moyenne (*MSE*) les trois approches, les performances d'estimation des trois méthodes sont différentes comme nous pouvons l'observer sur la figure 4.3. En effet nous constatons que très vite, la présence du pic pénalise totalement le *MLE3* avec des *MSEs* (en bleu) qui deviennent rapidement trop importantes, rendant les estimations de P_u , τ et swh totalement inexploitables.

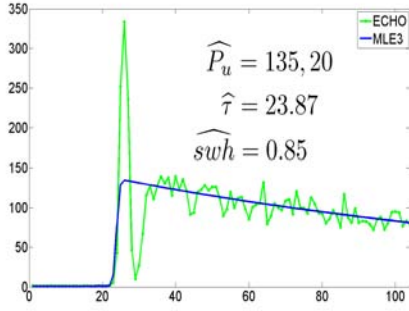
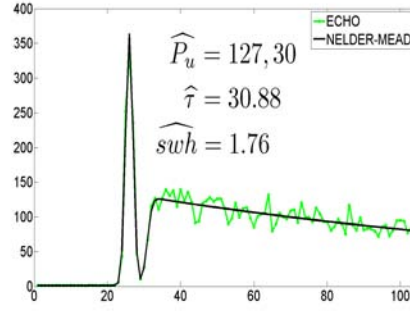
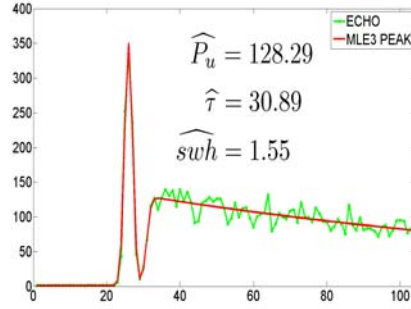
(a) Estimation par le *MLE3*.(b) Estimation par *Nelder-Mead*.(c) Estimation par le *MLE3 Peak*.

FIG. 4.2 – Échos estimés par les trois méthodes d'estimation (*MLE3*, *MLE3 Peak* et *Nelder-Mead*) d'une forme d'onde avec un pic avant le front de montée, $\theta_b = (P_u, \tau, sw) = (130, 31, 2)$ et $\theta_p = (A_1, k_1, v_1) = (400, 26, 1)$.

En regardant de plus près avec les zooms inclus dans chaque figure, nous nous rendons compte que le *MLE Peak* fournit des *MSEs* inférieures à celles obtenues avec l'estimateur *MLE3* mais supérieures à celles obtenues avec la méthode de *Nelder-Mead* dans le cas de l'estimation de τ et sw . De plus, pour ces deux derniers paramètres, l'écart en terme de *MSEs* entre le *MLE3 Peak* et *Nelder-Mead* s'accroît lorsque sw augmente. L'algorithme de *Nelder-Mead* permettant d'évaluer précisément la vraisemblance, *MLE3 Peak* diffère dans ces estimations à cause des approximations rappelées dans l'Annexe B.3. C'est en appréciant la différence des *MSEs*, principalement celles de sw , que nous pouvons

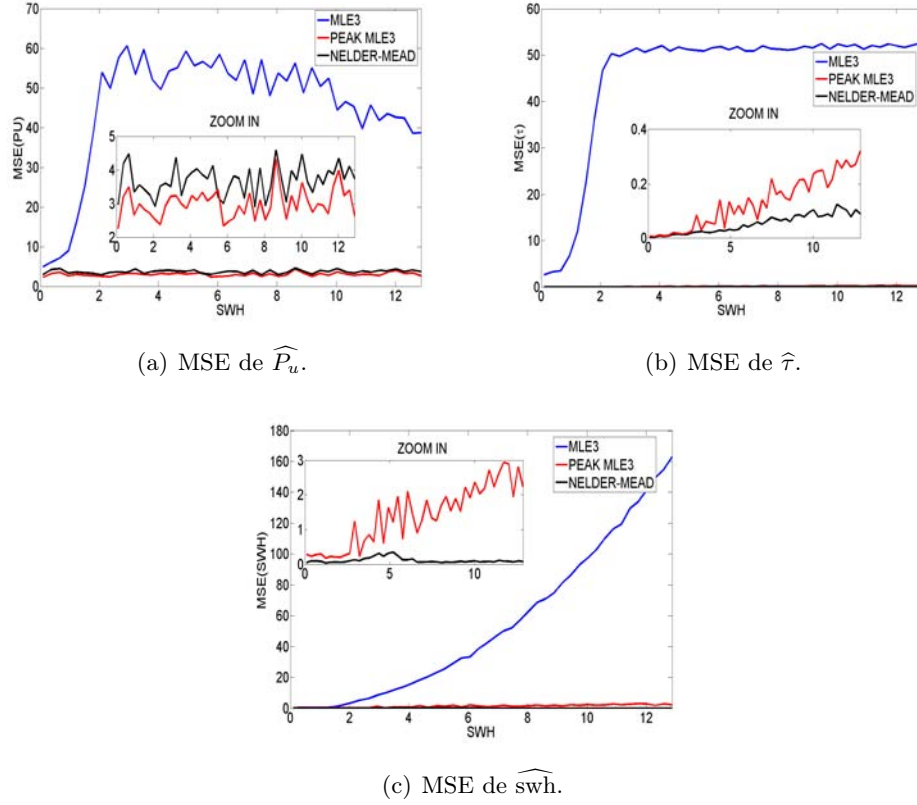


FIG. 4.3 – Comparaison des MSEs en fonction de swh dans le cas d'un pic avant le front de montée.

constater que *MLE3 Peak* n'est pas adapté aux signaux altimétriques de ce type. Si les résultats obtenus avec *Nelder-Mead* sont très satisfaisants, il ne faut pas oublier que son implantation est trop coûteuse en coût calculatoire. La version *MLE-Peak* que nous proposons permet d'obtenir des résultats satisfaisants pour ce type de signaux pour des hauteurs de vagues inférieures à $swh = 2.7\text{m}$, correspondant à des erreurs d'estimations de l'ordre de 55cm.

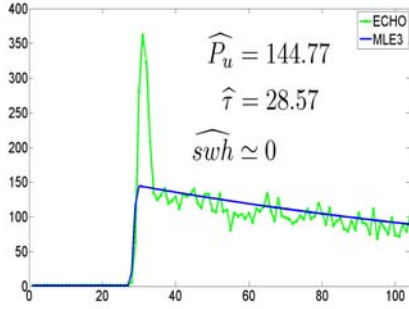
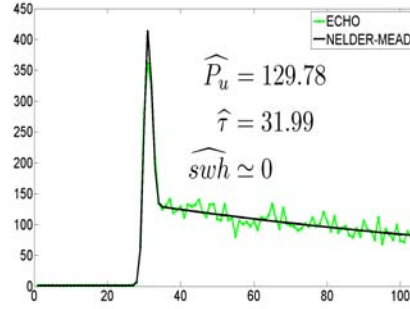
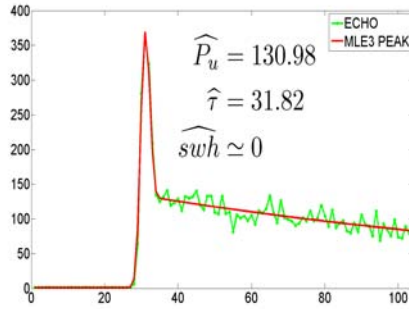
(a) Estimation par le *MLE3*.(b) Estimation par *Nelder-Mead*.(c) Estimation par le *MLE3 Peak*.

FIG. 4.4 – Échos estimés par les trois méthodes d'estimation (*MLE3*, *MLE3 Peak* et *Nelder-Mead*) d'une forme d'onde avec un pic sur le front de montée, $\theta_b = (P_u, \tau, swh) = (130, 31, 2)$ et $\theta_p = (A_1, k_1, v_1) = (400, 31, 1)$.

4.4.2 Pic sur le front de montée

Nous effectuons alors la même étude pour un pic situé au milieu du front de montée du signal altimétrique. La figure 4.4 montre un exemple d'estimation pour un écho dont les paramètres sont identiques à l'exemple de la figure 4.2 sauf la position du pic ($k_1 = 31$). On constate que le pic déforme totalement le front de montée et qu'il devient visuellement difficile de les distinguer. Lorsque l'on compare l'estimation *MLE3* (en bleu) et la forme d'onde bruitée (en vert) il est possible d'observer une surestimation de la puissance. En regardant plus précisément on se rend compte que les estimations de

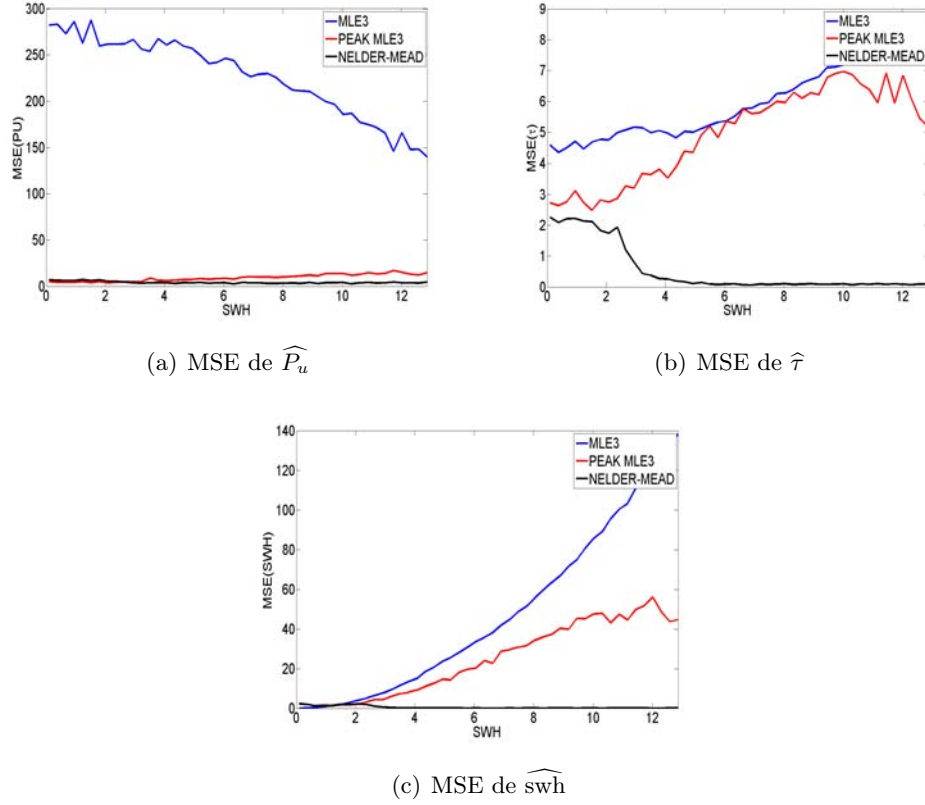


FIG. 4.5 – Comparaison des MSEs en fonction de swh dans le cas d'un pic sur le front de montée.

τ et swh faites par l'algorithme *MLE3* sont effectuées à partir du pic, et sont donc erronées. Nous pouvons constater d'ailleurs que tous les estimateurs conduisent à $swh = 0$ dans l'exemple de la figure 4.4. L'importance de cette déformation de l'écho se confirme en observant les performances d'estimation des algorithmes sur la figure 4.5. Comme précédemment les estimations *MLE3* fournissent des estimations incohérentes. Les résultats obtenus pour l'estimation de τ , illustrés sur la figure 4.5(b), montrent que pour des hauteurs de vagues moyennes comprises entre 0 et 3m, *MLE3 Peak* et *Nelder-Mead* obtiennent des estimations proches qui ne sont toutefois pas satisfaisantes. Ceci s'explique car pour des hauteurs de vagues faibles, la pente du front de montée (corrélée avec swh)

et le pic se confondent totalement. La figure 4.4 est un exemple de cette situation, nous permettant de nous rendre compte de la difficulté de les distinguer visuellement. Cependant, lorsque swh devient grand (figure 4.6) il devient plus aisé de faire la distinction. Ainsi, lorsque swh augmente, le nombre d'échantillons du signal sur le front de montée augmente, permettant à l'algorithme de *Nelder-Mead* de faire des estimations correctes. Mais dans ce cas, lorsque swh augmente, l'approximation faite dans le *MLE3 Peak* par rapport au maximum de vraisemblance exact devient de plus en plus sensible, rendant les résultats du *MLE3* non acceptables. Par comparaison avec la première catégorie de

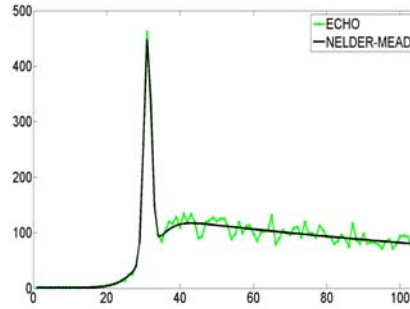


FIG. 4.6 – Forme d'onde avec $\text{swh} = 10\text{m}$ et un pic sur le front de montée.

pic étudiée, la présence d'un pic sur le front de montée est un problème plus compliqué. Notre algorithme *MLE3 Peak* n'est pas du tout adapté à ce type de problème, mais on se rend compte qu'il en est de même pour la méthode de référence *Nelder-Mead* pour l'estimation de τ lorsque $\text{swh} < 3\text{m}$.

4.4.3 Pic sur le plateau final

La dernière catégorie de pics regroupe les formes d'onde perturbées sur le plateau final comme nous l'illustrons sur la figure 4.7, avec $k_1 = 75$. Les deux algorithmes incluant les pics dans la vraisemblance obtiennent une bonne estimation de l'écho bruité. On remarque tout de même que l'approche *MLE3* surévalue la puissance du signal et fait une légère erreur sur le front de montée, illustrée sur la figure 4.7(a). Nous constatons

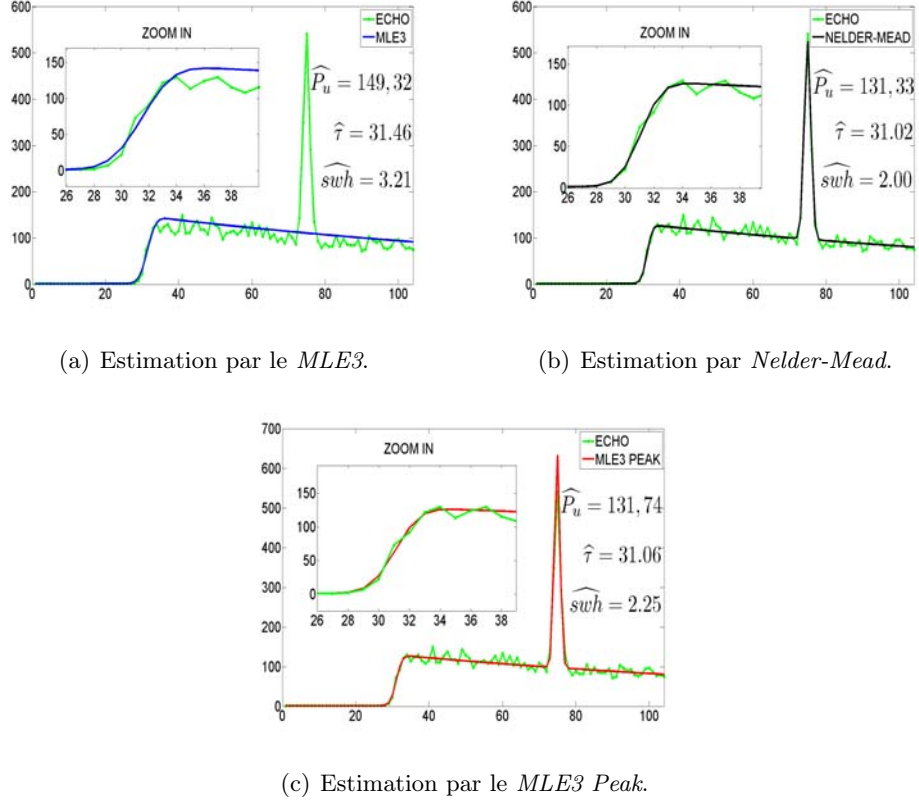


FIG. 4.7 – Échos estimés par les trois méthodes d'estimation (*MLE3*, *MLE3 Peak* et *Nelder-Mead*) d'une forme d'onde avec un pic sur le plateau final, $\theta_b = (P_u, \tau, swh) = (130, 31, 2)$ et $\theta_p = (A_1, k_1, v_1) = (400, 75, 1)$.

sur la figure 4.7(c) que *MLE3 Peak* permet d'effectuer une bonne estimation de l'écho dans sa globalité, et ce en accord avec le résultat de *Nelder-Mead* figure 4.7(b). La figure 4.8 met toujours en évidence la divergence rapide du *MLE3* mais il est aussi possible d'observer, en zoomant sur les *MSEs*, la similitude des performances entre le *MLE3 Peak* et *Nelder-Mead*, principalement pour les estimations de P_u et τ . Pour des hauteurs de vagues moyennes de plus de 12m, qui sont tout de même des événements très rares, la *MSE* sur l'estimation de swh en utilisant l'algorithme du *MLE3 Peak* n'excède pas les $1m^2$. Nous choisirons donc d'appliquer l'algorithme du *MLE3 Peak* plutôt sur ce type

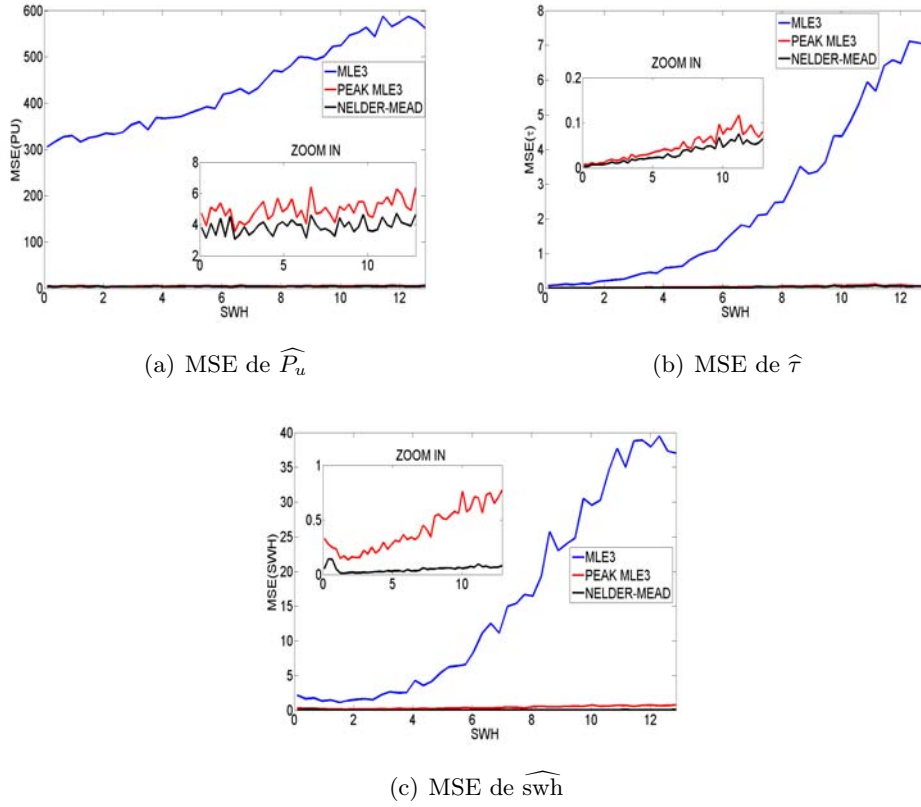


FIG. 4.8 – Comparaison des MSEs en fonction de swh dans le cas d'un pic sur le plateau final.

de signaux car l'implémentation numérique déjà évoquée permet tout de même d'obtenir des performances d'estimation appréciables.

4.4.4 Résultats sur des signaux réels

Les performances sur des signaux synthétiques étant encourageantes dans le cas d'un pic sur le plateau final, nous décidons de l'appliquer sur des signaux réels issus de la base de données PISTACH. Pour commencer, nous testons l'algorithme *MLE3 Peak* sur des signaux dit « Browniens » (non perturbés par un pic) afin de déterminer l'impact des modifications pour l'estimation de signaux de forme classique. La figure 4.9 montre les

estimations obtenues pour deux signaux différents. La première estimation illustrée sur la figure 4.9(a) met en avant le fait que les deux algorithmes, i.e *MLE3* et *MLE3 Peak*, peuvent obtenir des résultats très similaires. Dans ce cas, *MLE3 Peak* estime l'amplitude du pic égale à zéro. Cependant, la figure 4.9(b) nous permet de remarquer que le *MLE3 Peak* peut parfois interpréter une partie du bruit comme un pic. Toutefois, l'estimation de la partie « altimétrique » de l'écho reste semblable à celle obtenue avec le *MLE3*. Nous pouvons donc apprécier une forme de robustesse de l'estimateur *MLE3 Peak* qui ne détériore en rien l'estimation des paramètres P_u , τ et swh , même dans le cas de signaux de type « Brownien ». Nous observons ensuite la qualité des estimations du

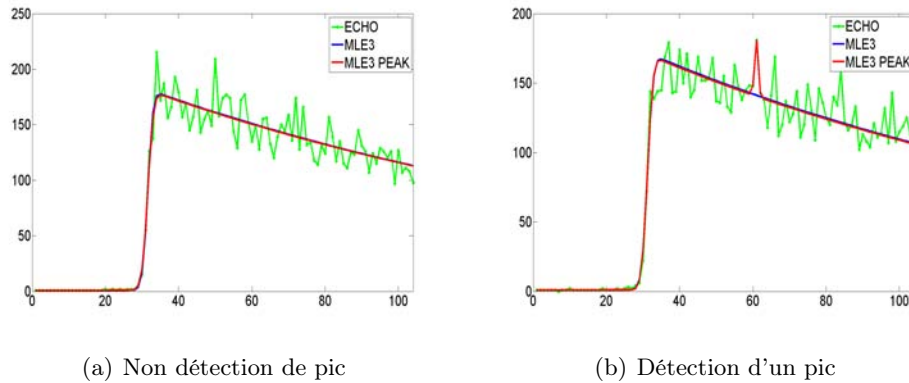
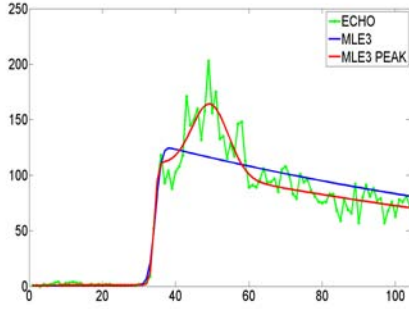
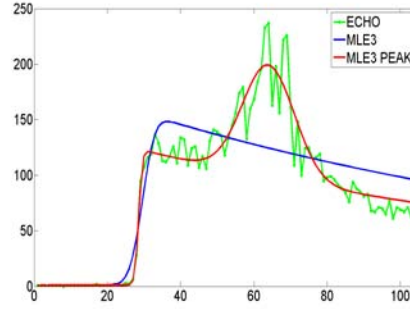


FIG. 4.9 – Comparaison des estimations de formes d'onde réelles de type « Brownien ».

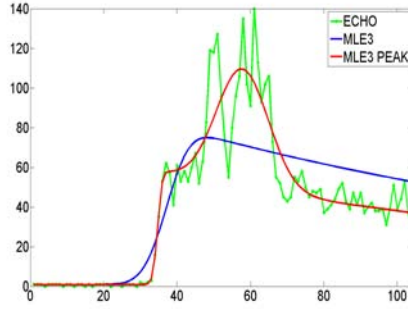
MLE3 Peak pour des signaux de la classe 21 du projet PISTACH (signaux perturbés par un ou plusieurs pics sur le plateau final). La figure 4.10 illustre trois exemples allant d'un ensemble de pics peu perturbateurs, figure 4.10(a), moyennement perturbateurs, figure 4.10(b) et très perturbateurs, figure 4.10(c). On remarque que l'importance des perturbations liées aux pics est en étroite corrélation avec les performances d'estimation obtenues avec la méthode *MLE3*. Plus la zone de perturbation est large et d'amplitude importante, plus les estimations du *MLE3* sont insatisfaisantes. À l'opposé, la méthode *MLE3 Peak* réussit parfaitement à estimer la partie « altimétrique » du signal ainsi que l'ensemble des perturbations considérées comme un pic. Il suffit d'observer comment le



(a) Pic de faible importance.



(b) Pic de moyenne importance.



(c) Pic de grande importance.

FIG. 4.10 – Comparaison des estimations de formes d'onde réelles avec perturbation d'un pic sur le plateau final.

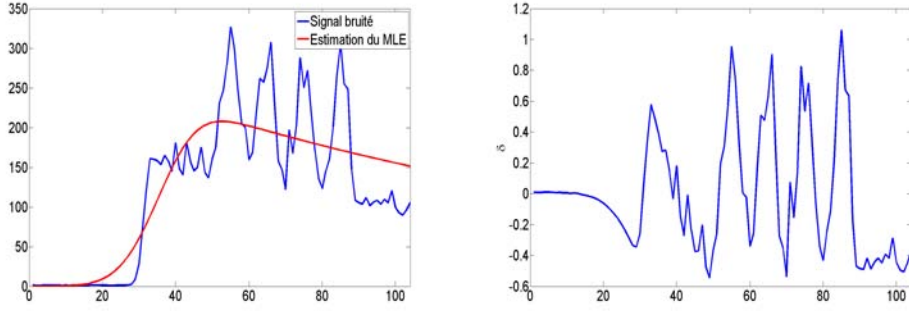
modèle *MLE3 Peak* suit bien le front de montée sur les trois figures 4.10(a), 4.10(b) et 4.10(c), à l'inverse du modèle de Hayne sur les figures 4.10(b) et 4.10(c), induisant une mauvaise estimation de τ et swh . La puissance P_u (qui correspond au maximum de l'écho de Brown estimé) est visiblement surestimée par l'estimateur *MLE3* dans les trois figures, alors que l'estimateur *MLE3 Peak* fournit une valeur qui paraît plus satisfaisante. Il est aussi possible de constater une certaine souplesse de l'algorithme sur la figure 4.10(c), car les perturbations pouvant être considérées comme deux pics distincts, sont ici estimées par *MLE3 Peak* comme un seul pic très large. Des résultats similaires sont présentés dans [TMST10a] en appliquant l'algorithme *MLE Peak* sur des signaux plus perturbés.

4.5 Extension à la présence de plusieurs pics

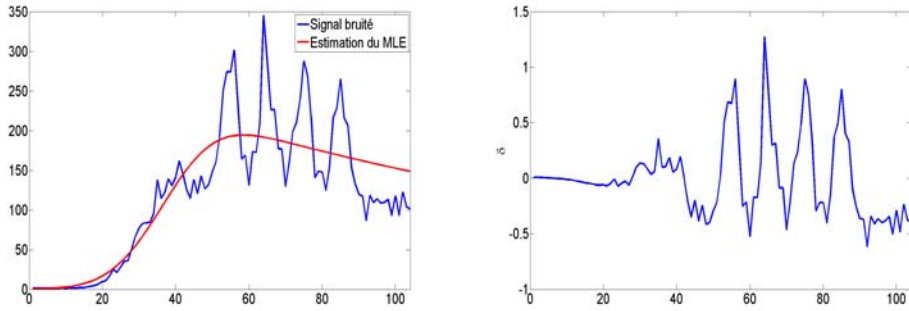
Après avoir validé l'intérêt de l'estimateur *MLE3 Peak* dans sa version « mono-pic », nous nous intéressons à sa version « multi-pics », dans laquelle $q > 1$. Le problème de l'estimation de q est un problème délicat qui mériterait une étude approfondie. Toutefois, dans une première approche, nous présentons ici une méthode heuristique permettant d'estimer q et de valider l'approche multi-pics. Nous considérons les pics comme des anomalies qui se superposent au modèle de Hayne (4.1). Partant de ce fait, nous décidons d'utiliser les résultats d'estimation du *MLE3* afin de détecter les différents pics à partir de :

$$\delta = \frac{\mathbf{y} - \mathbf{s}_{MLE3}}{\bar{y}}, \quad (4.14)$$

avec \mathbf{y} le signal bruité, \mathbf{s}_{MLE3} le modèle estimé par la méthode *MLE3* et \bar{y} la puissance moyenne du signal bruité. Les figures 4.11 mettent en avant deux exemples d'échos perturbés par 4 pics. Le premier exemple illustré figure 4.11(a) montre en bleu un écho synthétique généré avec une faible hauteur de vague moyenne (swh) et le modèle estimé par le *MLE3*. Les résultats de l'évaluation de δ s'observent sur la figure 4.11(b). Nous constatons 5 grandes variations de δ , les 4 plus importantes correspondant aux pics présents dans le signal. La première variation quant à elle est provoquée par la mauvaise estimation du front de montée par le *MLE3*. Nous constatons dans cet exemple qu'un seuil de détection compris en 0.6 et 0.8 peut être une solution. Dans le deuxième exemple illustré figure 4.11(c), l'écho est généré avec une hauteur de vague moyenne plus importante. La figure 4.11(d) nous permet d'observer l'évaluation de notre critère. Encore une fois, les 4 variations les plus importantes correspondent aux pics et l'utilisation d'un seuil au voisinage de $\delta = 0.8$ semble une solution adaptée. La détermination empirique de ce seuil a été confortée par de nombreuses simulations. Nous utilisons donc ce seuil sur δ pour détecter la présence de pics dans les formes d'onde analysées. Ainsi, un vecteur



(a) Écho perturbé par 4 pics avec un faible swh (b) δ obtenue pour l'écho avec un faible swh



(c) Écho perturbé par 4 pics avec un fort swh (d) δ obtenue pour l'écho avec un fort swh

FIG. 4.11 – Deux exemples d'échos perturbés par 4 pics et des δ associés.

indicateur de présence est créé :

$$\mathcal{I}_p(i) = \begin{cases} 0, & \text{si } \delta_i < 0.8, \\ 1, & \text{sinon,} \end{cases} \quad (4.15)$$

pour $i = (1, \dots, K)$. Ainsi, à partir de $\mathcal{I}_p(i)$ nous pouvons non seulement estimer le nombre de pics à prendre en considération mais aussi en initialiser les paramètres, en détectant les séquences de 1 que nous nommerons BB_j , et qui seront spécifiques à chaque pic détecté. Ainsi, $\forall j = (1, \dots, q)$, BB_j nous permet d'effectuer les initialisations suivantes pour le $j^{\text{ème}}$ pic :

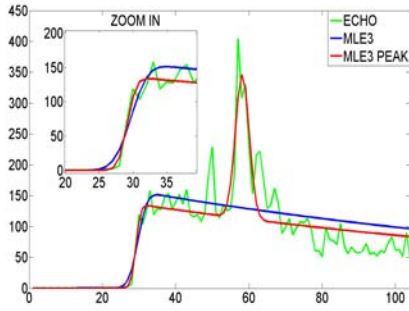
- la position k_j est initialisée à la porte centrale de BB_j ,

- l'amplitude s'initialise à partir du signal observé tel que : $A_j = y(k_j)$,
- enfin, la largeur du pic v_j correspond à la largeur du bloc BB_j .

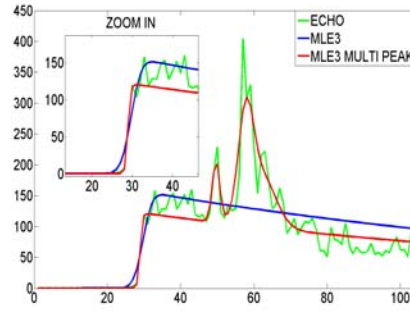
Comme précédemment, nous avons testé cette méthode de détection sur des échos de Brown synthétiques et réels. Sur la totalité des échos (synthétiques et réels) nous n'avons constaté aucune fausse détection. Contrairement aux résultats de la figure 4.9, cette nouvelle méthode de détection est moins sensible aux pics du bruit multiplicatif. De plus, dans certains cas, la méthode d'estimation à plusieurs pics permet d'améliorer les estimations des paramètres altimétriques en comparaison à l'estimateur *MLE3 Peak*. La figure 4.12 met en avant les résultats obtenus sur deux échos réels. Les figures 4.12(a) et 4.12(c) correspondent aux estimations obtenues en prenant compte un unique pic. Lorsqu'on observe les fronts de montée estimés (correspondant aux paramètres τ et swh) on se rend compte sur les figures 4.12(b) et 4.12(d) que l'approche multi-pics capte mieux le front de montée. Visuellement, nous voyons bien l'influence de l'algorithme multi-pics, et les différences d'estimation suggèrent des améliorations possibles pour l'estimation des pics. Nous pouvons donc en déduire qu'il est possible de rendre la méthode d'estimation des pics efficace sur un plus large spectre de signaux réels lorsque l'on considère $q > 1$. Quelques exemples d'applications sur des signaux réels sont illustrés dans [TSMT10].

4.6 Conclusion

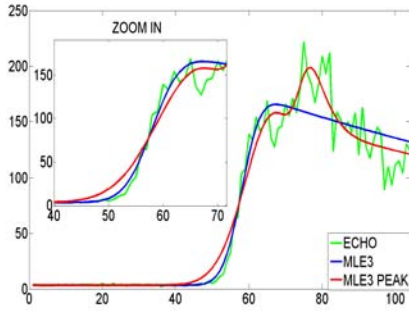
Suite aux résultats obtenus dans la partie consacrée à la classification de signaux altimétriques selon leurs formes nous avons proposé un algorithme permettant l'estimation des paramètres d'échos piqués. Tout d'abord nous nous sommes focalisés sur une méthode d'estimation tenant compte de la présence d'un seul pic. Après avoir analysé les performances selon la position du pic sur la forme d'onde, nous en avons conclu que cet estimateur est très performant lorsque le pic se situe sur le plateau final de la forme d'onde. Nous avons ensuite généralisé cette méthode en autorisant la présence de plusieurs pics. Ainsi, à partir d'une méthode heuristique de détection du nombre de pics présents dans



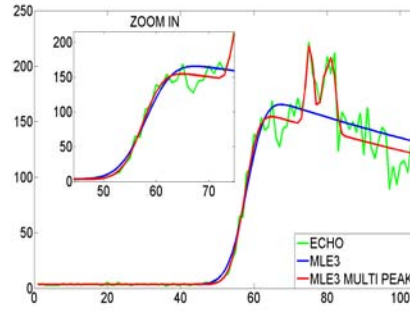
(a) Echo estimé avec la méthode mono-pic.



(b) Echo estimé avec la méthode multi-pics.



(c) Echo estimé avec la méthode mono-pic.



(d) Echo estimé avec la méthode multi-pics.

FIG. 4.12 – Comparaison des estimations entre la méthode mono-pic et l'approche multi-pics sur deux signaux réels.

l'écho, nous avons illustré sur des signaux réels l'intérêt d'une telle méthode et plus généralement l'intérêt du développement d'algorithmes spécifiques à certains types d'échos. L'exploitation des échos autres que « Browniens » référencés dans le projet PISTACH, figure 3.17, peut être simplifiée pour la mise en place d'algorithmes spécifiques comme ceux développés dans ce chapitre.

CHAPITRE 5

Conclusions et perspectives

Cette thèse a pour objet l'étude des signaux altimétriques. La succession des missions altimétriques de plus en plus précises permet la récolte d'un nombre croissant de données. Le traitement des données altimétriques nécessite donc une méthode d'estimation des échos très performante afin d'extraire le maximum d'informations des mesures. De plus, l'inhomogénéité de certaines mesures altimétriques se doit d'être traitée en amont, avant l'estimation, afin de classer les mesures par « type » ou par « forme » pour pouvoir par la suite les exploiter grâce à des algorithmes spécialisés. L'étude des côtes est notamment un des défis actuels puisque les signaux altimétriques associés ne peuvent être modélisés correctement ; cependant une solution existe.

Après une étude des bornes de Cramer-Rao de l'estimateur du maximum de vraisemblance actuellement utilisé pour l'estimation des signaux altimétriques [MTST08], nous avons défini une nouvelle méthode d'estimation bayésienne utilisant plusieurs types de lois *a priori* : des lois *a priori* non informatives, des lois *a priori* sur les vérités terrains ou encore des lois *a priori* dynamiques [SMTT08]. L'analyse des performances des estimateurs obtenus nous a permis de conclure que l'utilisation des lois *a priori* dynamiques est la méthode la plus pertinente. Elle permet en effet d'améliorer significativement la qualité des estimations de la hauteur de vague moyenne, mais aussi de définir un indice de confiance des estimations.

Dans un deuxième chapitre, nous nous intéressons à la classification des signaux altimétriques, tout d'abord par « type », afin d'identifier les échos de surfaces identifiées

(déserts, océans, eaux et forêts et glaces), puis par « forme », selon une table de classification issue du projet PISTACH [TMST10b]. L'ensemble des méthodes proposées pour la classification par « type » autorisent à partir d'un jeu de paramètres exclusivement altimétriques d'identifier clairement les différents types de surfaces analysées. De plus, une distinction entre les glaces du nord et les glaces du sud est rendue possible. Quant à la classification des signaux selon leur « forme », la chaîne de traitement proposée réussit à isoler les échos de « Brown », puis à identifier les signaux altimétriques selon leurs formes. Les performances de cette chaîne de traitement destinée à la classification sont satisfaisantes et prometteuses.

Enfin, dans un dernier chapitre nous avons proposé de modéliser les signaux altimétriques piqués [TMST10a], [TSMT10]. Pour cela nous les avons considérés comme des échos « Brownien » auxquels se superpose un ou plusieurs pics modélisés par une loi normale. La mise en place d'un nouvel estimateur et l'étude des performances d'estimations associées nous ont permis d'identifier l'utilité d'une telle approche dans le cas de pics situés sur le plateau final. Il est alors possible d'étendre l'estimation des signaux altimétriques à ce type de signaux souvent rencontrés près des côtes.

Différentes perspectives peuvent être dégagées à partir des travaux menés. Il conviendrait d'envisager d'étendre l'estimation bayésienne développée au premier chapitre de cette thèse aux futures missions altimétriques, et plus précisément à celles concernant l'interférométrie comme par exemple tout récemment CRYOSAT-2 ou encore, en 2016, la mission SWOT. Les mesures simultanées de ce type de missions permettraient en effet grâce au supplément d'information d'affiner la vraisemblance des données. Développer des algorithmes d'estimation [Hal13] et de classification dans ce contexte nous semble très prometteur.

Enfin, il serait intéressant d'approfondir les résultats du dernier chapitre. L'étude des signaux non-océaniques à proximité des côtes met en avant les performances satisfaisantes de l'estimateur *MLE3 Peak*, mais surtout celles de l'estimateur de *Nelder-Mead*. Or, les temps de calcul de cette méthode s'avèrent trop important pour rendre l'estimateur de *Nelder-Mead* véritablement opérationnel. Les résultats obtenus suggèrent cependant la possibilité de mettre en place une nouvelle méthode d'estimation plus précise que le *MLE3 Peak*, et plus rapide que la méthode de *Nelder-Mead*. Des résultats intéressants sont donc à attendre en utilisant d'autres méthodes d'estimation et éventuellement d'autres modèles de pics.

Annexes

ANNEXE A

Les bornes de Cramer-Rao

Soit $\mathbf{y} = y_1, \dots, y_K$ un écho altimétrique bruité et $\mathbf{s} = s_1, \dots, s_K$ le modèle à 3 paramètres (P_u , τ et swh) de Hayne (1.4) correspondant. Pour plus de lisibilité, nous noterons f la vraisemblance (1.9). La matrice d'information de Fisher pour le vecteur paramètre inconnu $\boldsymbol{\theta} = (P_u, \tau, \text{swh})$ est définie par :

$$\mathbf{F} = -E \begin{pmatrix} \frac{\partial^2 \ln f}{\partial P_u^2} & \frac{\partial^2 \ln f}{\partial P_u \partial \tau} & \frac{\partial^2 \ln f}{\partial P_u \partial \text{swh}} \\ \frac{\partial^2 \ln f}{\partial \tau \partial P_u} & \frac{\partial^2 \ln f}{\partial \tau^2} & \frac{\partial^2 \ln f}{\partial \tau \partial \text{swh}} \\ \frac{\partial^2 \ln f}{\partial \text{swh} \partial P_u} & \frac{\partial^2 \ln f}{\partial \text{swh} \partial \tau} & \frac{\partial^2 \ln f}{\partial \text{swh}^2} \end{pmatrix} \quad (\text{A.1})$$

avec $E[\cdot]$ l'espérance mathématique. Les variances de tous les estimateurs non biaisés des paramètres P_u , τ et swh sont caractérisées par une borne inférieure correspondant à la borne de Cramer-Rao [Kay93]. Cette borne de Cramer-Rao s'obtient en inversant la matrice de Fisher (A.1). Pour plus de simplicité, notons θ_i , $i=1, 2, 3$, les paramètres inconnus P_u , τ et swh . Tout estimateur $\hat{\theta}_i$ de θ_i vérifie l'inégalité suivante :

$$\text{Var}(\hat{\theta}_i) \geq \frac{-E\left(\frac{\partial^2 \ln f}{\partial \theta_j^2}\right) E\left(\frac{\partial^2 \ln f}{\partial \theta_k^2}\right) + \left[E\left(\frac{\partial^2 \ln f}{\partial \theta_j \partial \theta_k}\right)\right]^2}{\det(\mathbf{F})}, \quad i \neq j \neq k, \quad (\text{A.2})$$

avec $\det(\mathbf{F})$ le déterminant de \mathbf{F} . La partie de droite de l'inégalité est la borne de Cramer-Rao de θ_i . Le calcul de la borne de Cramer-Rao du paramètre θ_i se fait en déterminant l'espérance de la dérivée seconde de $\ln f$:

$$E\left(\frac{\partial^2 \ln f}{\partial \theta_j \partial \theta_k}\right) = L \sum_{n=0}^{K-1} \frac{-1}{x_n^2} \frac{\partial x_n}{\partial \theta_j} \frac{\partial x_n}{\partial \theta_k}, \quad (\text{A.3})$$

pour $i, j = 1, 2, 3$, $(\theta_i, \theta_j) \in \{P_u, \tau, \text{swh}\}^2$. Les éléments de la matrice d'information de Fisher (A.1) se calculent simplement, pour devenir :

$$\left\{ \begin{array}{l} F_{11} = \frac{LK}{P_u^2}, \\ F_{12} = \frac{L}{P_u} \sum_{k=0}^{K-1} G_{k,\tau,\text{swh}}, \\ F_{13} = \frac{L}{P_u} \frac{\text{swh}}{4c^2} \sum_{k=0}^{K-1} M_{k,\tau,\text{swh}}, \\ F_{22} = L \sum_{k=0}^{K-1} G_{k,\tau,\text{swh}}^2, \\ F_{23} = L \frac{\text{swh}}{4c^2} \sum_{k=0}^{K-1} G_{k,\tau,\text{swh}} M_{k,\tau,\text{swh}}, \\ F_{33} = L \frac{\text{swh}^2}{16c^4} \sum_{k=0}^{K-1} M_{k,\tau,\text{swh}}^2 \end{array} \right.$$

avec

$$\begin{aligned} G_{k,\tau,\text{swh}} &= \alpha - B_{\text{swh}} \frac{\exp\left(-A_{k,\tau,\text{swh}}^2\right)}{1 + \text{erf}\left(A_{k,\tau,\text{swh}}\right)} \\ M_{k,\tau,\text{swh}} &= \alpha^2 - D_{k,\tau,\text{swh}} \frac{\exp\left(-A_{k,\tau,\text{swh}}^2\right)}{1 + \text{erf}\left(A_{k,\tau,\text{swh}}\right)} \end{aligned}$$

et

$$\begin{aligned} A_{k,\tau,\text{swh}} &= \frac{kT_s - \tau - \frac{\alpha}{4c^2} \text{swh}^2 - \alpha \sigma_p^2}{\sqrt{2} \sqrt{\frac{\text{swh}^2}{4c^2} + \sigma_p^2}} \\ B_{\text{swh}} &= \sqrt{\frac{2}{\pi \left(\frac{\text{swh}^2}{4c^2} + \sigma_p^2\right)}} \\ D_{k,\tau,\text{swh}} &= \sqrt{\frac{2}{\pi}} \frac{kT_s - \tau + \frac{\alpha}{4c^2} \text{swh}^2 + \alpha \sigma_p^2}{\left(\frac{\text{swh}^2}{4c^2} + \sigma_p^2\right)^{3/2}}. \end{aligned}$$

Il est à noter que les fonctions $G_{k,\tau,\text{swh}}$ et $M_{k,\tau,\text{swh}}$ sont notées G_k et M_k pour plus de lisibilité. À partir de la matrice d'information de Fisher, nous pouvons alors déterminer les bornes de Cramer-Rao du maximum de vraisemblance :

$$\begin{aligned} BRC(P_u) &= \frac{P_u^2}{L} \frac{(\sum G_k^2)(\sum M_k^2) - (\sum G_k M_k)^2}{\Delta} \\ BRC(\tau) &= \frac{1}{L} \frac{K(\sum M_k^2) - (\sum M_k)^2}{\Delta} \\ BRC(\text{swh}) &= \frac{16c^4}{L\text{swh}^2} \frac{K(\sum G_k^2) - (\sum G_k)^2}{\Delta} \end{aligned} \quad (\text{A.4})$$

avec \sum représentant $\sum_{k=0}^{K-1}$, et :

$$\begin{aligned} \Delta = & K(\sum G_k^2)(\sum M_k^2) + 2(\sum G_k)(\sum M_k)(\sum G_k M_k) - (\sum G_k)^2(\sum M_k^2) \\ & - (\sum M_k)^2(\sum G_k^2) - K(\sum G_k M_k)^2. \end{aligned}$$

L'espérance mathématique des trois bornes de Cramer-Rao n'est pas facilement exploitable car G_k et M_k ont des expressions très complexes.

Il est à noter que nous n'étudions pas l'influence du nombre d'échantillons K sur les bornes car il s'agit d'un paramètre fixé lors de l'acquisition d'un écho. La figure A.1 met en avant le fait que les bornes de Cramer-Rao des paramètres sont des fonctions décroissantes en fonction du nombre de « look » L . Ce résultat n'est pas surprenant car le nombre de « look » influe directement sur le niveau de bruit du signal. De plus, nous pouvons exploiter cette figure afin de calculer les *MSEs* (Mean Square Error) et les bornes de Cramer-Rao du milieu du front de montée τ en mètre : dans ce cas, τ correspond à la distance entre le satellite et la surface de la mer que nous noterons d . En effet, τ en mètre peut aussi s'exprimer en secondes :

$$d = \frac{c\tau}{2},$$

d'où :

$$\text{MSE}(\hat{d}) = \frac{\text{MSE}(\hat{\tau}) c^2}{4},$$

avec \hat{d} et $\hat{\tau}$ les estimations de d et τ . Par exemple, pour $L = 90$, nous obtenons $\text{MSE}(\hat{\tau}) \approx -184\text{dB}$ donnant ainsi $\sqrt{\text{MSE}(\hat{d})} \approx 9.8\text{cm}$. Un calcul similaire permet

d'obtenir $\sqrt{BRC(d)} \approx 1.9\text{cm}$. Ainsi l'écart-type d'un estimateur optimal de la distance d est 5 fois inférieur à celui du MLE, correspondant à un gain de 8cm. Les figures A.2, A.3 et A.4 illustrent les bornes de Cramer-Rao des trois paramètres (P_u , τ et swh) en fonction de différents paramètres. La figure A.2 met en avant le fait que les bornes de Cramer-Rao de τ et swh ne dépendent pas de l'amplitude du signal P_u . La figure A.3 illustre la faible influence de τ sur l'ensemble des bornes de Cramer-Rao. Nous pouvons aussi constater sur la figure A.4 que la borne de Cramer-Rao de swh augmente lorsque swh augmente. Ainsi, toutes ces figures permettent de comparer les performances du *MLE* approché [Dum85] avec les *MSEs* optimales fournies par les bornes de Cramer-Rao. La totalité des *MSEs* ont été calculées avec 1000 réalisations de Monte Carlo de signaux synthétiques. Nous constatons que le *MLE* approché permet d'obtenir une estimation optimale de la puissance P_u . Cependant les différences entre les *MSEs* et les bornes de Cramer-Rao des paramètres τ et swh nous permettent de conclure qu'il est possible d'améliorer la qualité des estimations de ces deux paramètres.

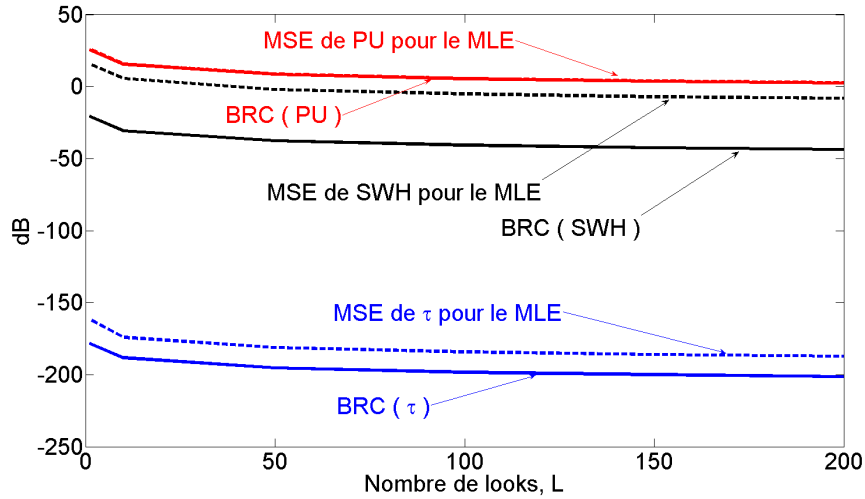


FIG. A.1 – Bornes de Cramer-Rao et *MSEs* du Pseudo-MLE pour les trois paramètres en fonction de L ($P_u = 160$, $\tau = 32$, swh = 6).

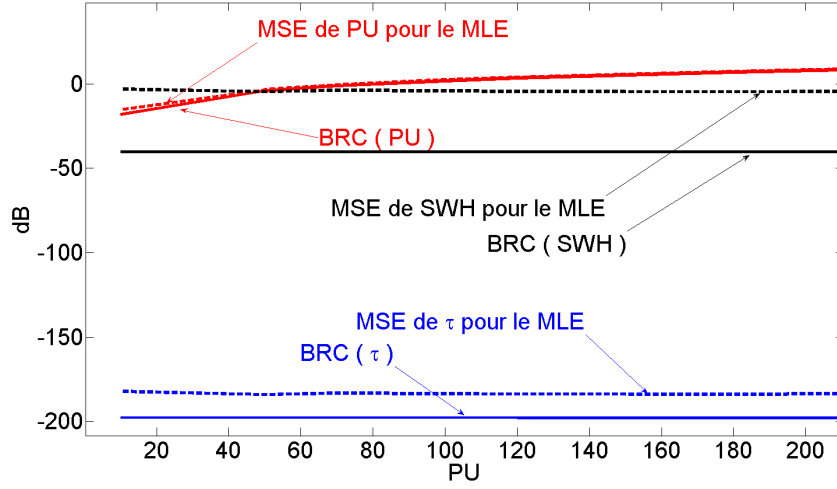


FIG. A.2 – Bornes de Cramer-Rao et $MSEs$ du Pseudo-MLE pour les trois paramètres en fonction de P_u ($L = 90$, $\tau = 32$, $swh = 6$).

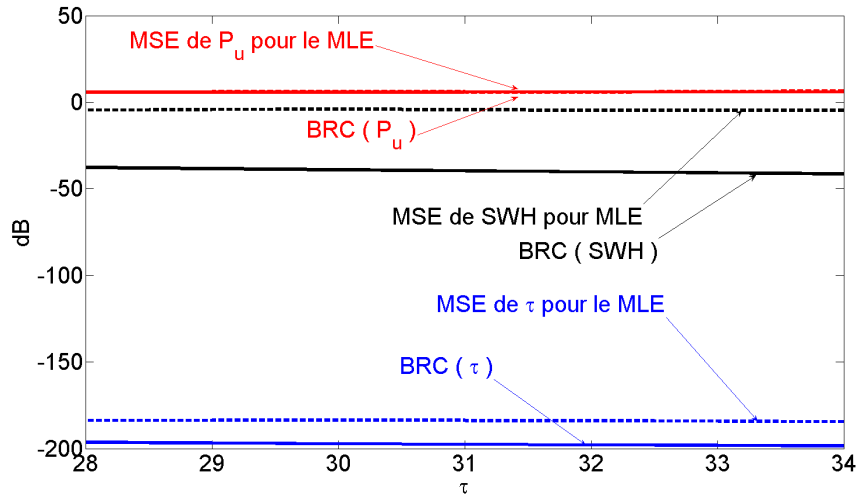


FIG. A.3 – Bornes de Cramer-Rao et $MSEs$ du Pseudo-MLE pour les trois paramètres en fonction de τ ($L = 90$, $swh = 6$, $P_u = 160$).

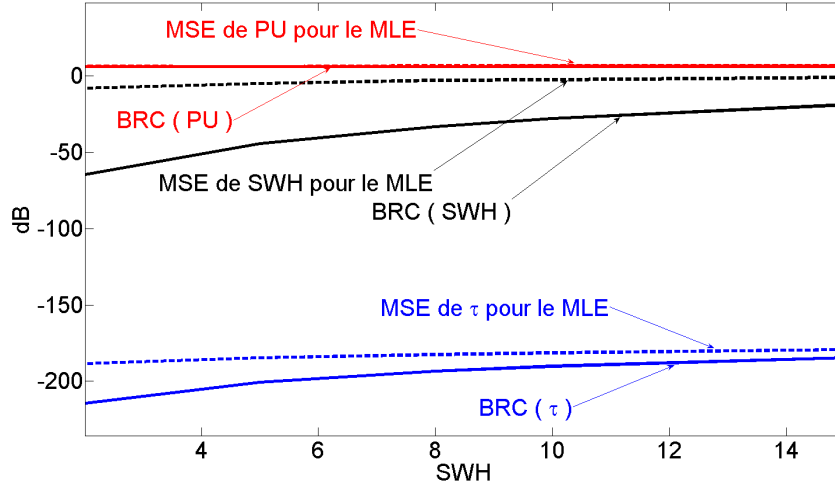


FIG. A.4 – Bornes de Cramer-Rao et $MSEs$ du Pseudo-MLE pour les trois paramètres en fonction de swh ($L = 90$, $\tau = 32$, $P_u = 160$).

Nous effectuons alors la même analyse en utilisant une modélisation à 4 paramètres (1.3). Le nouveau paramètre correspond à l'angle de dépointage de l'antenne (noté ξ) qui est principalement responsable du changement de pente du plateau final de la forme d'onde illustré figure 1.9. Plusieurs études ont montré l'utilité de l'estimation de ce paramètre [ATZ⁺04], particulièrement en présence de blooms ou de cellules de pluie. Ce quatrième paramètre dans la modélisation des données apporte quelques changements au modèle, plus précisément :

$$\begin{aligned} P_u &\rightarrow P_u \exp\left(-\frac{4}{\gamma} \sin^2 \xi\right), \\ \alpha &\rightarrow \alpha \left[\cos(2\xi) - \frac{\sin^2(2\xi)}{\gamma} \right]. \end{aligned} \quad (\text{A.5})$$

Comme cela a été fait précédemment, la matrice d'information de Fisher est une matrice 4×4 dont les éléments se définissent par :

$$F_{ij} = -E\left(\frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_j}\right), \quad (\text{A.6})$$

pour $i, j = 1, \dots, 4$ avec $\theta_i \in \{P_u, \tau, \text{swh}, \xi\}$. Dans ce cas à 4 paramètres nous utilisons la notation suivante :

$$V_{k,\tau,\text{swh},\xi} = I_{k,\tau,\text{swh},\xi} + B_{\text{swh}} J_{\text{swh},\xi} \frac{\exp\left(-A_{k,\tau,\text{swh},\xi}^2\right)}{1 + \operatorname{erf}\left(A_{k,\tau,\text{swh},\xi}\right)} \quad (\text{A.7})$$

avec

$$I_{k,\tau,\text{swh},\xi} = -2 \sin(2\xi) \left[\left(1 + \frac{2}{\gamma} \cos(2\xi)\right) \left(-\alpha k + \alpha \tau\right) + \left(\alpha^2 \sigma_p^2 + \frac{\alpha^2 \text{swh}^2}{4c^2}\right) \left(\cos(2\xi) - \frac{\sin^2(2\xi)}{\gamma}\right) + \frac{2}{\gamma} \right],$$

et

$$J_{\text{swh},\xi} = 2 \sin(2\xi) \left(1 - \frac{2}{\gamma} \cos(2\xi)\right) \left(\alpha \sigma_p^2 + \frac{\alpha \text{swh}^2}{4c^2}\right).$$

Les éléments de la matrice d'information de Fisher relatifs à ξ sont de la forme :

$$\left\{ \begin{array}{l} F_{14} = \frac{L}{P_u} \sum_{k=0}^{K-1} V_{k,\tau,\text{swh},\xi}, \\ F_{24} = L \sum_{k=0}^{N-1} G_{k,\tau,\text{swh},\xi} V_{k,\tau,H,\xi}, \\ F_{34} = L \frac{\text{swh}}{4c^2} \sum_{k=0}^{N-1} M_{k,\tau,\text{swh},\xi} V_{k,\tau,\text{swh},\xi}, \\ F_{44} = L \sum_{k=0}^{N-1} V_{k,\tau,\text{swh},\xi}^2. \end{array} \right.$$

Les autres éléments de la matrice d'information de Fisher F_{ij} , $(i, j) \in \{1, 2, 3\}^2$ ont les mêmes expressions (A), à l'exception que α est remplacé à partir de l'égalité A.5. Ce remplacement est aussi valable pour les fonctions $A_{k,\tau,\text{swh}}$, $D_{k,\tau,\text{swh}}$, $G_{k,\tau,\text{swh}}$ et $M_{k,\tau,\text{swh}}$ qui deviennent alors $A_{k,\tau,\text{swh},\xi}$, $D_{k,\tau,\text{swh},\xi}$, $G_{k,\tau,\text{swh},\xi}$ et $M_{k,\tau,\text{swh},\xi}$. La figure A.5 illustre la perte des performances pour une estimation à 4 paramètres à la place de 3. Par exemple pour $L = 90$ nous avons $\sqrt{\text{MSE}(\hat{d})} \approx 21.2\text{cm}$ and $\sqrt{BRC(d)} \approx 2.7\text{cm}$ à comparer avec

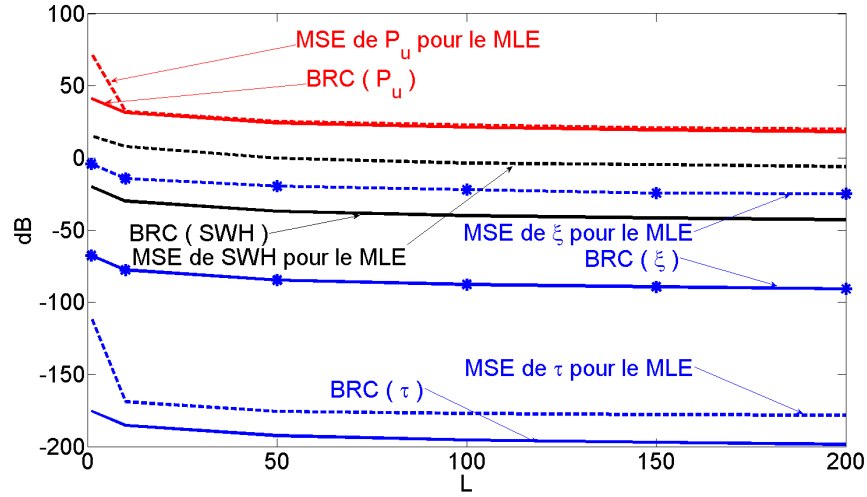


FIG. A.5 – Bornes de Cramer-Rao et $MSEs$ du Pseudo-MLE pour les quatre paramètres en fonction de L ($P_u = 160$, $\tau = 32$, $swh = 6$, $\xi = 0.1$).

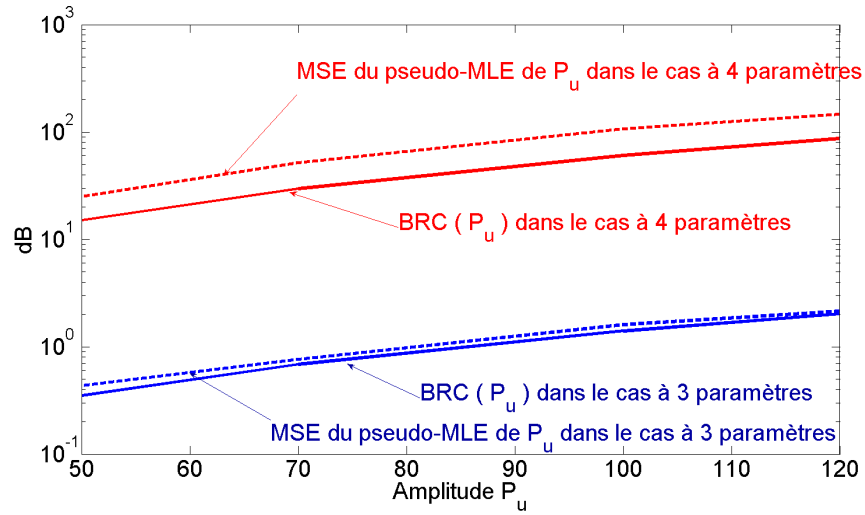


FIG. A.6 – Comparaison entre les estimations de P_u avec un modèle de Brown à 3 et 4 paramètres ($L = 90$, $\tau = 32$, $swh = 6$, $\xi = 0.1$).

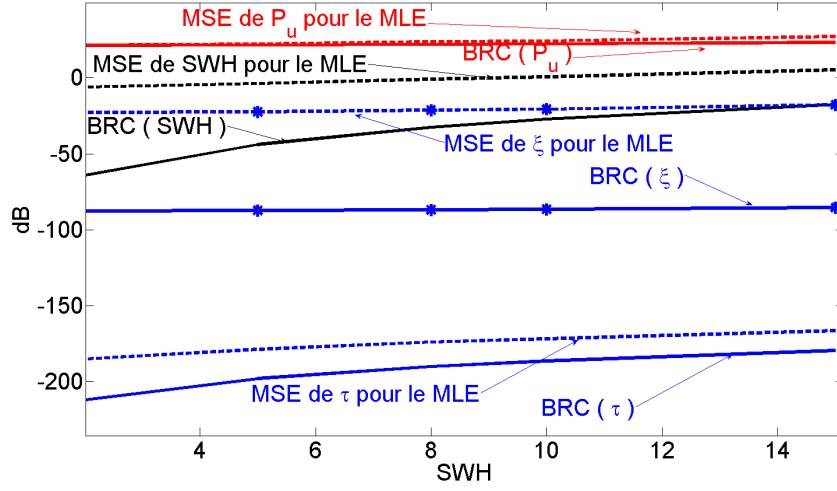


FIG. A.7 – Bornes de Cramer-Rao et $MSEs$ du Pseudo-MLE pour les quatre paramètres en fonction de swh ($L = 90$, $\tau = 32$, $P_u = 160$, $\xi = 0.1$).

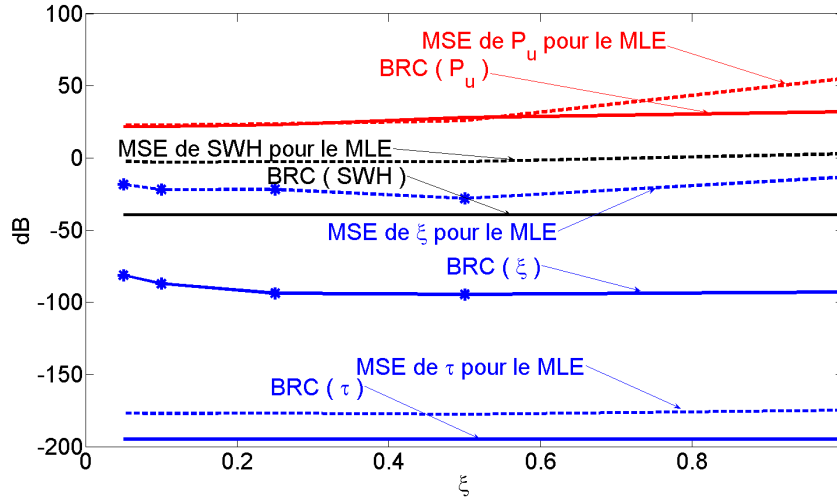


FIG. A.8 – Bornes de Cramer-Rao et $MSEs$ du Pseudo-MLE pour les quatre paramètres en fonction de ξ ($L = 90$, $\tau = 32$, $P_u = 160$, $swh = 6$).

le résultat précédent. La perte de précision de l'estimation pour le paramètre P_u lorsque nous passons à un modèle à 4 paramètres s'observe sur la figure A.6. En particulier nous remarquons que l'estimation de P_u en utilisant le *MLE* approché ne reste pas une estimation optimale dans le cas d'un modèle à 4 paramètres. Les figures A.7 et A.8 représentent les bornes de Cramer-Rao des 4 paramètres en fonction de la hauteur de vague moyenne sw_h et en fonction du dépointage de l'antenne ξ . Il est intéressant de noter que le comportement des différentes bornes de Cramer-Rao (BRC) est comparable à celui obtenus avec une modélisation à 3 paramètres, c'est-à-dire non optimale. Cependant, l'estimation de P_u semble moins précise pour de grandes valeurs de sw_h et ξ . L'ensemble de ces performances renforce l'idée de l'étude d'un nouvel estimateur afin d'augmenter la précision des estimations.

ANNEXE B

Méthode de Quasi-Newton

Soit $\mathbf{y} = y_1, \dots, y_K$ un écho altimétrique bruité, l'estimateur du maximum de vraisemblance permet de déterminer le vecteur $\boldsymbol{\theta}_b$ décrivant au mieux les données observées \mathbf{y} en utilisant le modèle de Hayne s_k . Selon la modélisation utilisée nous avons $\boldsymbol{\theta}_b = (P_u, \tau, \text{swh})^T$, $\boldsymbol{\theta}_b = (P_u, \tau, \text{swh}, \xi)^T$ ou encore $\boldsymbol{\theta}_b = (P_u, \tau, \text{swh}, \xi, \lambda_s)^T$. En nous positionnant dans le cas général, nous caractérisons le vecteur $\boldsymbol{\theta}_b$ comme un vecteur à m composantes. Autrement dit, il s'agit d'estimer le vecteur $\boldsymbol{\theta}_b$ qui maximise la vraisemblance f des données ou la log-vraisemblance [PCT90] :

$$\ln [f(\mathbf{y}|\boldsymbol{\theta}_b)] = \sum_{k=1}^K \ln \left[\frac{L^{KL}}{(L-1)!} \right] + (L-1) \ln(y_k) - L \ln(s_k) - L \frac{y_k}{s_k}. \quad (\text{B.1})$$

avec L le nombre de *look*, s_k le modèle de Hayne, \mathbf{y} l'écho observé et K le nombre de portes. La méthode actuellement utilisée pour estimer ce vecteur $\boldsymbol{\theta}_b$ inverse le problème en le transformant en un problème de minimisation d'une fonction coût C :

$$C = -\ln [f(\mathbf{y}|\boldsymbol{\theta}_b)]. \quad (\text{B.2})$$

L'approche utilisée pour estimer le vecteur paramètre trouve son origine dans la méthode d'optimisation sans contrainte de Quasi-Newton [NW06]. Définissant un vecteur $\boldsymbol{\theta}_{b0}$ d'initialisation, l'algorithme de Quasi-Newton est une méthode itérative de descente de gradient, telle que le vecteur minimisant la fonction coût C est mis à jour à chaque itération de la manière suivante [NW06] :

$$\boldsymbol{\theta}_{b(t+1)} = \boldsymbol{\theta}_{b(t)} - \Psi H_{(t)}^{-1} \nabla C_{(t)}, \quad (\text{B.3})$$

avec :

- Ψ un coefficient compris entre 0 et 1 (tout de même proche de 1), il intervient dans la vitesse de convergence de l'algorithme.
- $H_{(t)}$, une matrice symétrique et inversible. Il s'agit de la hessienne de la fonction coût C , plus précisément la matrice des dérivées secondes à l'instant (t) .
- Enfin, ∇ est l'opérateur du gradient.

L'expression du gradient de la fonction coût C se calcule à partir de l'expression de la log-vraisemblance (B.1) pour un vecteur paramètre $\boldsymbol{\theta}_b = (\theta_{b,1}, \dots, \theta_{b,m})$, m étant le nombre de paramètres composant le vecteur :

$$\nabla C = \begin{pmatrix} \frac{-\partial \ln[f(\mathbf{y}|\boldsymbol{\theta}_b)]}{\partial \theta_{b,1}} \\ \vdots \\ \frac{-\partial \ln[f(\mathbf{y}|\boldsymbol{\theta}_b)]}{\partial \theta_{b,m}} \end{pmatrix} = L \begin{pmatrix} \sum_{k=1}^K \frac{s_k - y_k}{s_k^2} \frac{\partial s_k}{\partial \theta_{b,1}} \\ \vdots \\ \sum_{k=1}^K \frac{s_k - y_k}{s_k^2} \frac{\partial s_k}{\partial \theta_{b,m}} \end{pmatrix}. \quad (\text{B.4})$$

Soit $\mathbf{D} = (D_k)_{k=1, \dots, K}$ et $\mathbf{B} = (B_{ik})_{i=1, \dots, m, k=1, \dots, K}$ tels que :

$$D_k = \frac{s_k - y_k}{s_k}, \quad (\text{B.5})$$

$$B_{ik} = \frac{\partial s_k}{s_k \partial \theta_{b,i}}, \quad (\text{B.6})$$

ainsi le gradient de la fonction coût C s'exprime sous la forme du produit matriciel suivant :

$$\nabla C = L \mathbf{B} \mathbf{D}^T. \quad (\text{B.7})$$

De la même manière nous développons l'expression de la hessienne de la fonction coût C dont les termes génériques s'écrivent :

$$\begin{aligned}
H_{ij} &= \frac{\partial^2 C}{\partial \theta_{b,i} \partial \theta_{b,j}}, \\
&= L \frac{\partial}{\partial \theta_{b,i}} \left[\sum_{k=1}^K \frac{s_k - y_k}{s_k^2} \frac{\partial s_k}{\partial \theta_{b,j}} \right], \\
&= L \sum_{k=1}^K \frac{\partial}{\partial \theta_{b,i}} \left(\frac{s_k - y_k}{s_k^2} \right) \frac{\partial s_k}{\partial \theta_{b,j}} + \frac{s_k - y_k}{s_k^2} \frac{\partial^2 h_k}{\partial \theta_{b,i} \partial \theta_{b,j}}.
\end{aligned}$$

C'est à ce niveau-là qu'intervient l'approximation faite dans les algorithmes actuellement utilisés qui ne sont plus alors de véritables estimateurs du maximum de vraisemblance mais seulement une version *approchée*. En effet, on suppose : $\forall k \in [1, K], s_k - y_k \simeq 0$. Il est ainsi possible de négliger le dernier terme de l'expression :

$$\begin{aligned}
H_{ij} &= L \sum_{k=1}^K \frac{\partial}{\partial \theta_{b,i}} \left(\frac{s_k - y_k}{s_k^2} \right) \frac{\partial s_k}{\partial \theta_{b,j}}, \\
&= L \sum_{k=1}^K \frac{1}{s_k^4} \left[s_k^2 \frac{\partial s_k}{\partial \theta_{b,i}} - 2(s_k - y_k) s_k \frac{\partial s_k}{\partial \theta_{b,i}} \right] \frac{\partial s_k}{\partial \theta_{b,j}}.
\end{aligned}$$

L'hypothèse sur $s_k - y_k \simeq 0$ nous permet de la même manière de négliger les termes croisés des dérivées secondes pour aboutir à :

$$\begin{aligned}
H_{ij} &= L \sum_{k=1}^K \frac{1}{s_k^2} \frac{\partial s_k}{\partial \theta_{b,i}} \frac{\partial s_k}{\partial \theta_{b,j}}, \\
&= L \sum_{k=1}^K B_{ik} B_{jk}.
\end{aligned}$$

Finalement, l'expression de l'approximation de la hessienne est :

$$\mathbf{H} = L \mathbf{B} \mathbf{B}^T. \quad (\text{B.8})$$

La solution itérative de la méthode d'estimation en utilisant l'algorithme de Quasi-Newton s'écrit [Dum85] :

$$\boldsymbol{\theta}_{b(t+1)} = \boldsymbol{\theta}_{b(t)} - \Psi(\mathbf{B} \mathbf{B}^T)^{-1} \mathbf{B} \mathbf{D}^T. \quad (\text{B.9})$$

Un simple calcul des dérivées partielles permet donc d'obtenir un estimateur *approché* du maximum de vraisemblance dont le biais et la variance sont étudiés dans [Dum85]. Les dérivées partielles des paramètres P_u , τ et $\text{sw} \ y$ sont décrites plus en détail. La méthode du maximum de vraisemblance actuellement utilisée pour estimer les échos altimétriques correspond à cette approche, à ceci près qu'une simplification supplémentaire intervient dans les équations (B.5) et (B.6) pour devenir :

$$D_k = \frac{s_k - y_k}{P}, \quad (\text{B.10})$$

$$B_{ik} = \frac{\partial s_k}{P \partial \theta_{b,i}}, \quad (\text{B.11})$$

P étant une constante.

Liste des publications

1. Jérôme Severini, Corinne Mailhes, Pierre Thibaut, and Jean-Yves Tournet. Bayesian estimation of altimeter echo parameters. In *IEEE International Geoscience & Remote Sensing Symposium, Boston, USA, 06/07/08- 11/07/08*. IEEE, juillet 2008.
2. Corinne Mailhes, Jean-Yves Tournet, Jérôme Severini, and Pierre Thibaut. Cramer-Rao bounds for radar altimeter waveforms. In *European Signal and Image Processing Conference (EUSIPCO), Lausanne (Suisse), 25/08/2008-29/08/2008*, page (electronic medium), [http ://www.eurasip.org/](http://www.eurasip.org/), août 2008. EURASIP.
3. Jean-Yves Tournet, Corinne Mailhes, Jérôme Severini, and Pierre Thibaut. Parameter estimation for peaky altimetric waveforms. In *Oceans from Space, Venise, Italie, 26/04/10-30/04/10*, 2010.
4. Jean-Yves Tournet, Corinne Mailhes, Jérôme Severini, and Pierre Thibaut. Shape Classification of altimetric signals using anomaly detection and Bayes decision rule (regular paper). In *IEEE International Geoscience & Remote Sensing Symposium, Honolulu, USA, 25/07/2010-30/07/2010*. IEEE, 2010.
5. P. Thibaut, J. Severini, C. Mailhes, and J-Y. Tournet. A multi-peak model for peaky altimetric waveforms. In *4th Coastal Altimetry Workshop, Porto, Portugal, 14/10/10-15/10/10*, 2010.

Bibliographie

- [All06] T. Allen. The story of gander. *Sensors*, 6(3) :249–259, 2006.
- [ATZ⁺04] L. Amarouche, P. Thibaut, O. Z. Zanife, J.-P. Dumont, P. Vincent, and N. Steunou. Improving the Jason-1 ground retracking to better account for attitude effects. *Marine Geodesy*, 27 :171–197, 2004.
- [BG98] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4) :434–455, Dec. 1998.
- [Boe06] J-R De Boer. Estimation des paramètres altimétriques par réseau de neurones. Master’s thesis, Université de Toulouse - ENSEEIHT, 2006.
- [Bou08] J. Bouffard. *Amélioration de l’altimétrie côtière appliquée à l’étude de la circulation dans la partie nord du bassin occidental méditerranéen*. PhD thesis, Université de Toulouse, Toulouse, FRANCE, 2008.
- [Bro77] G.S. Brown. The average impulse response of a rough surface and its applications. *IEEE Trans. Antennas and Propagation*, 25 :67–74, Jan. 1977.
- [CES⁺06] V. Rosmorduc (CLS), J. Benveniste (ESA), O. Lauret (Silogic), M. Milagro (SERCO), and N. Picot (CNES). *Radar Altimetry Tutorial*, 2006.
- [CZH01] Y. Chen, X. Zhou, and T. S. Huang. One-class svm for learning in image retrieval. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, pages 34–37 vol.1, 2001.

- [DFL⁺10] M. Durand, Lee-Lueng Fu, D.P. Lettenmaier, D.E. Alsdorf, E. Rodriguez, and D. Esteban-Fernandez. The surface water and ocean topography mission : Observing terrestrial surface water and oceanic submesoscale eddies. *Proceedings of the IEEE*, 98(5) :766 –779, may 2010.
- [DGL96] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer Verlag, New York, NY, 1996.
- [DH73] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. Wiley Interscience, 1973.
- [DT07] M. Davy and J.Y. Tournieret. Classification bayésienne supervisée par processus de dirichlet. *Actes du Colloque GRETSI*, 2007.
- [Dum85] J.-P. Dumont. *Estimation optimale des paramètres altimétriques des signaux radar Poséidon*. PhD thesis, Université de Toulouse - ENSEEIHT, Toulouse, FRANCE, 1985.
- [Fra02] C.R. Francis. Design of the cryosat system. In *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, volume 3, pages 1759 – 1761 vol.3, 24-28 2002.
- [Gar99] A. Garcia. *Study of the Origins of the Sigma-0 Blooms*. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1999.
- [GCSR95] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian data analysis*. Chapman & Hall, 1995.
- [GEVQ⁺ar] J. Gómez-Enri, S. Vignudelli, G.D. Quartly, C.P. Gommenginger, P. Cipolini, P.G. Challenor, and J. Benveniste. Modeling envisat RA-2 waveforms in the coastal zone : Case-study of calm water contamination. *IEEE Geosci. and Remote Sensing Letters*, to appear.
- [GR07] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *IEEE Trans. Signal Processing*, 7(4) :457–551, April 2007.

- [Hal13] A. Halimi. *Estimation des paramètres des signaux altimétriques*. PhD thesis, ENSEEIHT, Toulouse, FRANCE, 2010-2013.
- [Hay80] G.S. Hayne. Radar altimeter mean return waveforms from near normal incidence ocean surface scattering. *IEEE Trans. Antennas and Propagation*, 28(5) :687–692, Sept. 1980.
- [JZZ07] B. Jiang, M. Q. Zhang, and X. Zhang. Oscar : One-class SVM for accurate recognition of cis-elements. *Bioinformatics*, pages btm473+, October 2007.
- [Kay93] S. Kay. *Fundamentals of statistical signal processing : estimation theory*. Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [LRWW98] J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright. Convergence properties of the nelder mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9 :112–147, 1998.
- [Mah36] P. C. Mahalanobis. On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, volume 2, pages 49–55, April 1936.
- [MPRM⁺08] C. Martin-Puig, G. Ruffini, J. Marquez, D. Cotton, M. Srokosz, P. Challenor, K. Raney, and J. Benveniste. Theoretical model of sar altimeter over water surfaces. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 3, pages III –242 –III –245, 7-11 2008.
- [MTST08] Corinne Mailhes, Jean-Yves Tournet, Jérôme Severini, and Pierre Thibaut. Cramer-Rao bounds for radar altimeter waveforms. In *European Signal and Image Processing Conference (EUSIPCO), Lausanne (Suisse), 25/08/2008-29/08/2008*, page (electronic medium), <http://www.eurasip.org/>, août 2008. EURASIP.
- [NCK02] A. Elisseeff N. Cristianini, J. Shawe-Taylor and J. Kandola. On kernel-target alignment. In *Advances in Neural Information Processing Systems 14*, volume 14, pages 367–373, 2002.

- [NM65] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7 :308–313, 1965.
- [NW06] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, 2006.
- [Oll06] A. Ollivier. *Nouvelle approche pour l'extraction de paramètres géophysiques des mesures en altimétrie radar*. PhD thesis, Institut National Polytechnique de Grenoble, Grenoble, FRANCE, 2006.
- [PCT90] M.A. Srokosz P.G. Challenor and R.T. Tokmakian. Maximum likelihood estimation for radar altimetry. In *Proc. IEE Colloquium on Monitoring the Sea*, pages 10/1–10/3, London, UK, Dec. 1990.
- [Pot07] J-B. Pothin. *Techniques de sélection et d'élaboration de noyaux adaptés*. PhD thesis, Université de technologie de Troyes - UTT, Troyes, FRANCE, 2007.
- [QGS96] G. D. Quartly, T. H. Guymer, and M. A. Srokosz. The effects of rain on topex radar altimeter data. *Journal of Atmospheric and Oceanic Technology*, 13(6) :1209–1229, 1996.
- [QJR⁺07] A. Quesney, E. Jeansou, C. Ruiz, N. Steunou, B. Cugny, N. Picot, J.C. Souyris, S. Thiria, and M. Lebbah. Unsupervised classification of altimetric waveform over all surface type. *Proc. Ocean Surface Topography Science Team (OSTST)*, 2007.
- [QSG98] G. D. Quartly, M. A. Srokosz, and T. H. Guymer. Understanding the effects of rain on radar altimeter waveforms. *Advances in Space Research*, 22(11) :1567 – 1570, 1998. Remote sensing : Earth, ocean and atmosphere.
- [Qua97] G. Quartly. Achieving accurate altimetry across storms : Improved wind and wave estimates from C band. *Journal of Atmospheric and Oceanic Technology*, 14(3) :705–715, 1997.

- [RC89] E. Rodriguez and B. Chapman. Extracting ocean surface information from altimeter returns : The deconvoluion method. *Journal of Geophysical Research*, 94 :9761–9778, July 1989.
- [RC99] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 1999.
- [RD07] S. Rossignol and M. Davy. Détection de ruptures à l’aide des n SVM 1 classe z pour la segmentation des signaux sonores musicaux. In *GRETSI Groupe d’Étude du Traitement du Signal et des Images*, pages 165–168, septembre 2007.
- [RLV99] F. Remy, B. Legresy, and P. Vincent. New scientific opportunities from ka-band altimetry. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS ’99 Proceedings. IEEE 1999 International*, volume 1, pages 506–508 vol.1, 1999.
- [RR98] C.P. Robert and S. Richardson. *Discretization and MCMC Convergence Assessment*. Springer-Verlag, 1998.
- [Sap90] G. Saporta. *Probabilités, Analyse des Données et Statistique*. Technip, Paris, France, 1990.
- [SMTT08] Jérôme Severini, Corinne Mailhes, Pierre Thibaut, and Jean-Yves Tourneret. Bayesian estimation of altimeter echo parameters. In *IEEE International Geoscience & Remote Sensing Symposium, Boston, USA, 06/07/08-11/07/08*. IEEE, juillet 2008.
- [Sol05] J. Solle. Classification de données altimétriques radar. Master’s thesis, Université de Toulouse - ENSEEIHT, 2005.
- [SPT⁺99] B. Schölkopf, J. C. Platt, J. S. Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. Technical report, Microsoft Research, 1999.

- [TCRV06] J. Tournadre, B. Chapron, N. Reul, and D. C. Vandemark. A satellite altimeter model for ocean slick detection. *J. Geophys Res.*, 2006.
- [TD04] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Mach. Learn.*, 54(1) :45–66, 2004.
- [TFF07] P. Thibaut, F. Ferreira, and P. Femenias. Sigma-0 blooms in the envisat radar altimeter data. *Envisat Symposium*, 2007.
- [Thi08] P. Thibaut. Rapport sur la classification des formes d’onde dans PISTACH. Technical report, Collecte Localisation Satellites, 2008.
- [TMS10] J.Y. Tourneret, C. Mailhes, and J. Severini. Amélioration des produits altimétriques hauturiers pour les missions futures. Technical report, ENSEEIHT-IRIT-TéSA, 2010.
- [TMST10a] Jean-Yves Tourneret, Corinne Mailhes, Jérôme Severini, and Pierre Thibaut. Parameter estimation for peaky altimetric waveforms. In *Oceans from Space, Venice, Italie, 26/04/10-30/04/10*, 2010.
- [TMST10b] Jean-Yves Tourneret, Corinne Mailhes, Jérôme Severini, and Pierre Thibaut. Shape Classification of altimetric signals using anomaly detection and Bayes decision rule (regular paper). In *IEEE International Geoscience & Remote Sensing Symposium, Honolulu, USA, 25/07/2010-30/07/2010*. IEEE, 2010.
- [TSMT10] P. Thibaut, J. Severini, C. Mailhes, and J-Y. Tourneret. A multi-peak model for peaky altimetric waveforms. In *4th Coastal Altimetry Workshop, Porto, Portugal, 14/10/10-15/10/10*, 2010.
- [VVB⁺00] P. Vincent, J. Verron, P. Bahurel, P.Y. Le Traon, L. Eymard, F. Remy, J.F. Cretaux, J. Tournadre, and B. Chapron. The altika altimetry mission : science objectives and system definition. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International*, volume 7, pages 3036 –3038 vol.7, 2000.

- [ZDZ08] O. Zammit, X. Descombes, and J. Zerubia. Unsupervised one-class SVM using a watershed algorithm and hysteresis thresholding to detect burnt areas. In *Pattern Recognition and Image Analysis (PRIA)*, Nijni Novgorod Russie, 2008.

